Vol. II 1931

THE ANNALS

of

MATHEMATICAL STATISTICS

(Printed in U. S. A.)

Published and Lithoprinted by
EDWARDS BROTHERS, INC.
ANN ARBOR, MICH.

EDITORIAL COMMITTEE

A quarterly publication of the American Statistical Association, devoted to the theory and application of Mathematical Statistics.

Six dollars per annum.

Reprints of any article in this volume may be obtained at any time from the Editor at the following rates, postage included.

Number of Copies	Cost per Page
1- 4 . .	2 cents
5-24 . .	1½ cents
50 and over .	¾ cent
25-49 . .	1 cent

ADDRESS: Editor, Annals of Mathematical Statistics
Post Office Box 171, Ann Arbor, Michigan

THE ANNALS OF MATHEMATICAL STATISTICS

VOL. II FEBRUARY NO. 1

CONTENTS

PUBLISHED QUARTERLY BY
AMERICAN STATISTICAL ASSSOCIATION

Publication Office—Edwards Brothers, Inc., Ann Arbor, Mich.
Editorial Office—University of Michigan, Ann Arbor, Mich.

Application for entry as second class matter pending

EDITORIAL COMMITTEE

A quarterly publication sponsored by the American Statistical Association, devoted to the theory and application of Mathematical Statistics.

Six dollars per annum

Reprints of any article in this issue may be obtained at any time from the Editor at the following rates, postage included.

Number of copies	Cost per page
1- 4	2 cents
5-24	1½ cents
25-49	1 cent
50 and over	¾ cent

ADDRESS: Editor, Annals of Mathematical Statistics
Post Office Box 171, Ann Arbor, Michigan

THE RELATIONS BETWEEN STABILITY AND HOMOGENEITY*

By

L. v. Bortkiewicz

The idea of investigating the stability of statistical frequencies from the standpoint of the theory of probability goes back to the French mathematician Bienaymé. From various examples taken from social and moral statistics, he was the first to establish the fact that, almost without exception, the stability in question was essentially less than the "classical norm," that is, less than the expectation which is associated with the classical scheme of independent trials with a constant underlying probability. In order to explain this discrepancy between theory and observation, Bienaymé used a modification of the traditional procedure which was characterized by the assumption that between neighboring trials in a time ordered sequence a sort of dependence existed. Though interesting in itself and among other things adopted by Cournot as his own, we shall replace this method in what follows by another, originating from Lexis, which has the advantage of a wider usefulness, in that it can be applied not only

---

*Translated by A. R. Crathorne. Read before the American Statistical Association at Cleveland, Ohio, December 30, 1930.

to undulatory but to evolutory sequences.[1]

Let us assume that for a series of z successive time intervals, say years, we have found that some event (accident, death, marriage, crime) has happened $x_1, x_2, \ldots$ times, and that the corresponding number of "trials," that is the numbers of persons observed, are $s_1, s_2, \ldots$ so that the quotients $y_1 = \frac{x_1}{s_1}, y_2 = \frac{x_2}{s_2}, \ldots$ represent a time ordered sequence of relative frequencies. Instead of assuming, as the traditional theory demands, that each term y_k of this series corresponded to a common fundamental probability p, weighted with accidental errors, Lexis assumed that each value y_k was associated with a distinct probability p_k.

As a result of this, the expected amplitude of the fluctuations of the values y_k increased, and the greater the variations in the p_k's the greater the amplitude. Under the simplifying hypothesis s_k = const. ($= s$), the corresponding standard deviation σ is defined by

$$\sigma^2 = \frac{1}{z}\sum_{k=1}^{z}(y_k - y)^2, \quad y = \frac{1}{z}\sum_{k=1}^{z} y_k$$

For the case of a constant p we may write

$$E(\sigma^2) = \frac{z-1}{z} \cdot \frac{p(p-1)}{s} \tag{1}$$

where E denotes "expectation." In the Lexis procedure with a variable p_k, using the notation

$$\frac{z-1}{z} \cdot \frac{p(1-p)}{s} = u^2, \quad \frac{1}{z}\sum_{k=1}^{z} p_k = p, \quad p_k - p = \varepsilon_k, \quad \frac{1}{z}\sum_{k=1}^{z} \varepsilon_k^2 = \omega^2,$$

---

[1]Bienayme, in the journal "L'Institute," Vol. 7 (1831), pages 187-189, and in "Journal de la Societe de Statistique de Paris," 17e (1876), pages 199-204.
A. Cournot, Exposition de la theorie des chances et des probabilities, Paris, 1843, Nos. 79 and 117.
W. Lexis, "Uber die Theorie der Stabilitat statistischer Reihen," in the Jahrbuch fur Nationalokonomie und Statistik, Vol. 32 (1879), pages 60 . ., reprinted in Abhandlungen zur Theorie der Bevölkesungs und Moralstatistik, Jena, 1903, pages 170-212.

the corresponding relation

$$E(\sigma^2) = u^2 + \frac{zs - z + 1}{zs}\,\omega^2 \tag{2}$$

can be derived.[1]

In the following numerical examples the numbers of observations s_k are never less than some ten thousands, while $z = 10$. Hence, as far as these and similar examples are concerned, the numerical results are not appreciably altered if, instead of (2), we use

$$E(\sigma^2) = u^2 + \omega^2 \tag{3}$$

However, a certain inaccuracy arises, if, in the application of formula (3) to the raw data, one has disregarded the fundamental assumption that s_k is constant and in the expression for u^2 has replaced s by the arithmetic mean of the z values s_k. If, however, the latter differ little from one another, such a procedure gives rise to no great discrepancy. Lexis called the quantities u and ω in formula (3) the two "fluctuation components," which combine (according to the law of composition of forces) to give the expected total fluctuation. The quantity u gives expression to the effect of the "accidental causes" in the sense of the theory of probability, and this effect grows less and less with increasing s until it vanishes for $s = \infty$. For this reason Lexis called u the normal component. He also used the term "unessential fluctuation component." On the other hand, ω depends on the variations of the fundamental probability, that is on the underlying general conditions, and in this sense was designated by Lexis as the physical component. We may also

[1] One does not find formula (2) in Lexis's work. He was satisfied at this point with a rather inexact method yielding an approximate result. However, this did not affect the essential part of his discussion.

call it the essential component.

The first of the two components u and ω can be easily calculated directly with sufficient approximation. The usual method is to substitute for the unknown p in the expression for u^2 the value y, the arithmetic mean of the frequencies y_k, obtaining

$$u^2 = \frac{z-1}{z} \cdot \frac{y(1-y)}{s} \tag{4}$$

As for the second component ω, it is calculated by the indirect method of substituting σ^2 for $E(\sigma^2)$ in (3) and then ω is found from $\omega^2 = \sigma^2 - u^2$. This method, however, assumes that $\sigma > u$, or what is the same thing, that the dispersion coefficient, $Q = \frac{\sigma}{u}$, is greater than 1. In his older papers, Lexis distinguished between subnormal, normal and supernormal dispersion, according to whether Q was distinctly less than 1, approximately equal to 1, or distinctly greater than 1, and found that in social and moral statistics the subnormal dispersion never occurred and the normal rarely. Supernormal dispersion was the rule. So Lexis based his scheme of a varying underlying probability on the case of supernormal dispersion. In fact, from formula (3), we have

$$E(Q^2) = 1 + \left(\frac{\omega}{u}\right)^2 \tag{5}$$

which says that the variations in the underlying probability lead us to expect values of Q greater than unity.[1]

Notwithstanding the fact that Q was usually greater than unity, Lexis did not consider this a proof that his scheme ade-

[1]Under the influence of accidental causes, Q may be less than unity not only for constant, but also for varying underlying probabilities, and this circumstance must be considered in the determination of ω. It would carry us too far afield to go further into this matter.

quately described the actual facts. In addition to this he was more concerned with the fact that in experience Q showed a tendency to decrease with decreasing number of "trials," that is with decreasing s. Indeed, in a series of examples, Lexis had shown that a value of Q which was decidedly greater than unity when calculated for an entire country, decreased to nearly 1 when the data for the single administration districts of the same country were used. Lexis considered such behavior of Q as entirely in harmony with his scheme.

If we write formula (5) in the form

$$(6) \qquad E(Q^2) = 1 + s\,\frac{z\omega^2}{(z-1)p(1-p)},$$

we see that the excess of Q^2 over and above 1 is in expectation directly proportional to s. This was the explanation of the decrease of Q with decreasing s, for as Lexis said, we have no ground to expect that s being large or small had any bearing on the value of ω.

It is this last point about which the criticism of Lexis's dispersion theory centers. Notwithstanding the endeavors of Lexis to fit his theory to statistical reality, we can show that the facts were against him as far as his assumption that ω is fundamentally independent of s is concerned. If this assumption were true, then formula (6) tells us distinctly how Q decreases with diminishing s. We learn from experience that as a rule this decrease in Q is less than that given by the formula; from which it follows that the essential component, ω, has a tendency to increase with decreasing s.

If we desire to investigate just what happens in reality, a certain complication arises, because we are never able to compare groups which differ among one another as to s, but not as to p (or y). In order to eliminate to some extent the variations of p we consider the ratio of ω to p. Let $\frac{\omega}{p} = \beta$,

and call β the *relative* essential component to distinguish it from the *absolute* essential component ω . Formula (6) then becomes the following:

$$E(Q^2) = 1 + sp \frac{z\beta^2}{(z-1)(1-p)} \tag{7}$$

The product sp can be considered as the expected number of "successes." For a constant s_k ($= s$) we have

$$E(x_k) = sp_k, \qquad E\left(\frac{1}{z}\sum_{k=1}^{z} x_k\right) = sp$$

and, letting $s = \frac{1}{z}\sum s_k$, the last relation is true with sufficient approximation for a variable s_k provided the variation is not too pronounced. Let $sp = m$. Often, as in the examples which follow, p is so small that we can consider ($1-p$) as equal to 1. Formula (7) then becomes

$$E(Q^2) = 1 + \frac{z}{z-1} m\beta^2 \tag{8}$$

The question as to whether there is a connection between s and ω is now changed to an investigation of the relationship between m and β. In undertaking such an investigation empirically, we compare as to the behavior of m and β a statistical aggregate considered as a total with its component parts considered as partial aggregates. Let the number of the partial aggregates be n , and let the corresponding values of m and β as well as u , ω and σ be indicated by the subscript i , which can also serve as the ordinal number of the partial aggregate. For the total aggregate, let $i = 0$. The symbols $s_{i,k}$, $x_{i,k}$, $y_{i,k}$, $p_{i,k}$, are the s . x . y , p of the i th partial aggregate and the k th time interval. We also use

the notation

$$s_i = \frac{1}{z}\sum_{k=1}^{z} s_{i,k}\,, \quad x_i = \frac{1}{z}\sum_{k=1}^{z} x_{i,k}\,,$$

$$y_i = \frac{1}{z}\sum_{k=1}^{z} y_{i,k}\,, \quad p_i = \frac{1}{z}\sum_{k=1}^{z} p_{i,k}\,.$$

from which we have

$$s_0 = \sum_{i=1}^{n} s_i\,, \qquad x_0 = \sum_{i=1}^{n} x_i\,,$$

$$y_0 = \frac{1}{s_0}\sum_{i=1}^{n} s_i y_i\,, \quad p_0 = \frac{1}{s_0}\sum_{i=1}^{n} s_i p_i\,.$$

We have also the following relations:

$$m_i = s_i p_i\,, \quad \sigma_i^2 = \frac{1}{z}\sum_{k=1}^{z}(y_{i,k} - y_i)^2,$$

$$u_i^2 = \frac{z-1}{z} \cdot \frac{p_i(1-p_i)}{s_i}\,, \quad \omega_i^2 = \frac{1}{z}\sum_{k=1}^{z} e_{i,k}^2$$

$$\text{where} \quad e_{i,k} = p_{i,k} - p_i\,, \quad \beta_i = \frac{\omega_i}{p_i}\,,$$

$$E(\sigma_i^2) = u_i^2 + \omega_i^2\,, \quad Q_i = \frac{\sigma_i}{u_i}\,,$$

and using the notation $\frac{e_{i,k}}{p_{i,k}} = \varepsilon_{i,k}$ we have further

$$\beta_i^2 = \frac{1}{z}\sum_{k=1}^{z} \varepsilon_{i,k}^2$$

Finally, corresponding to formula (8), we have

$$E(Q_i^2) = 1 + \frac{z}{z-1} m_i \beta_i^2 \tag{9}$$

We shall now apply these formulas to statistics on the frequency of suicides in Germany for the decade 1902-1911. The numbers of "trials," $s_{i,k}$, are here the populations of the regions in question; the "successes," $x_{i,k}$, are the numbers of suicides for each year. The relative frequencies, $y_{i,k}$, are found by dividing the numbers of suicides by the corresponding populations. Like various other kinds of social phenomena, the suicides in pre-war German statistics were grouped according to states, the provinces of Prussia, right Rhenish Bavaria and left Rhenish Bavaria being included as states. In this way we have forty territories of very unequal size. For the decade 1902-1911, the mean population of the territories ranged from a maximum of 6,587,000 (Rhine Province) to a minimum of 45,000 (Schaumburg-Lippe). The maximum average number of suicides per annum was 1453 (Saxony) and the minimum 7 (Schaumburg-Lippe). Corresponding to the purpose of the investigation, these suicide figures x_i, which can be considered as approximations to m_i, were arranged in descending order, with $x_1 = 1453$ and $x_{40} = 7$.

For the whole of Germany, we have $x_0 = 13173$, y_0 $214 \cdot 10^{-6}$ (that is an average number of 214 suicides per annum for each million population). The ten values $y_{0,k}$ vary between $204 \cdot 10^{-6}$ and $223 \cdot 10^{-6}$. These fluctuations are markedly greater than one expects from the classical norm. The calculation of the dispersion-quotient gives $Q_0 = 3.14$, and, as the Lexis theory demands, is greater than any one of the 40 values of Q_i.[1] These values give 2.03 as a maximum and 0.75 as a

[1]A study of suicides and of homicides in the United States yields much the same general results as those shown here for suicides in Germany. (Note by the translator.)

minimum. Fixing attention on the eight smallest values of x_i, we find an average value of 1.02 for Q_i, and of the eight values, three are larger and five less than 1. So in this example the dispersion becomes very nearly 1 by narrowing the observation field.

But we have still to find out whether Q_i decreases with x_i according to the measure of decrease that one would expect under the hypothesis that β_i is fundamentally independent of x_i. To decide this question, we let β_i = const. = β, including $\beta_0 = \beta$, and substitute also x_i for m_i in formula (9). We have then on the one hand in expected values

$$Q_0^2 = 1 + \frac{z}{z-1} x_0 \beta^2$$

and on the other hand

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1 + \frac{z}{z-1} \cdot \frac{x_0}{n} \beta^2$$

from which follows

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1 + \frac{1}{n}(Q_0^2 - 1)$$

However, in our example, we find

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1.56, \qquad 1 + \frac{1}{n}(Q_0^2 - 1) = 1.22$$

and the difference 0.34 cannot be ascribed to chance for it is three times the probable error (the determination of which we cannot now take up). We must, then, assume that the average of the values β_i, for $i = 1$ to 40, is greater than β_0. Why this is so we shall see in the following discussion.

We consider now the mutual relationship between the deviations $\varepsilon_{i,k}$ and $\varepsilon_{j,k}$ which refer to two arbitrary territories N_i and N_j, and we build up according to the formula for a

correlation coefficient the expression

$$\gamma_{i,j} = \frac{1}{z} \sum_{k=1}^{z} \frac{\varepsilon_{i,k}\,\varepsilon_{i,k}}{\beta_i\,\beta_j}$$

The number of combinations of the subscripts i and j is $\frac{n(n-1)}{2}$, so there are that many values $\gamma_{i,j}$. Finally we construct a weighted arithmetic mean of these values according to the formula,

$$\gamma = \frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} m_i\, m_j\, \beta_i\, \beta_j\, \gamma_{i,j}}{\sum_{i=1}^{n} \sum_{j=i+1}^{n} m_i\, m_j\, \beta_i\, \beta_j}$$

The expression γ serves to characterize the mutual relationship of time ordered series of fundamental probabilities $p_{i,k}$, hence also of relative frequencies $y_{i,k}$, which may be considered as approximations to $p_{i,k}$. If we give the name "syndromy" to such an array of simultaneously distinct fundamental probabilities (or relative frequencies), we may call γ a "coefficient of syndromy." For $\gamma = 1$, we shall speak of "isodromy," for $1 > \gamma > 0$, of "homodromy," for $\gamma = 0$, of "paradromy," and for $\gamma < 0$, of "antidromy." We may include the last three cases, namely $\gamma < 1$, under the name "anisodromy."

With the help of γ we can exhibit the relation between β_0 on the one hand and the n values $\beta_1, \beta_2, \ldots \beta_n$ on the other hand as follows:

$$m_0^2\, \beta_0^2 = \sum_{i=1}^{n} m_i\, \beta_i^2 + \gamma \left\{ \left(\sum_{i=1}^{n} m_i\, \beta_i \right)^2 - \sum_{i=1}^{n} m_i^2\, \beta_i^2 \right\} \tag{10}$$

Since $m_\bullet = \sum_{i=1}^{n} m_i$, we find for $\gamma = 1$, from (10)

$$\beta_\circ = \frac{\sum_{i=1}^{n} m_i \beta_i}{\sum_{i=1}^{n} m_i} \tag{11}$$

and for $\gamma < 1$

$$\beta_\circ < \frac{\sum_{i=1}^{n} m_i \beta_i}{\sum_{i=1}^{n} m_i} \tag{12}$$

Hence, only in the case of isodromy is the assumption justified that the relative essential fluctuation component for the total aggregate is as large as that for the partial aggregates. In every other case, namely for anisodromy, the relative essential component for the total aggregate falls below the level for the partial aggregates more and more as γ becomes less and less.

In the suicide example under consideration we have homodromy, which is reasonable, since the fluctuations in suicide frequency in the single states are influenced in part by factors which are not local but general for all Germany. Somewhat tedious calculations give $\gamma = 0.38$. At the same time we find $\beta_\bullet = 0.0246$ approximately, while the average for β_i, $i = 1$ to 40 is 0.0392.

If now we group the 40 states into five groups so that states numbered 1 to 8 form the first group, states numbered 9 to 16 the second, and so on, we find as average values of β_i, 0.0354, 0.0358, 0.0485, 0.0528 and 0.0767. The quantities β_i then show a tendency to increase as x_i (or m_i) decreases.

If, as in this example, the total aggregate is a "natural unit," we should expect to have homodromy in the vast majority of cases. On the other hand, we should expect paradromy if the total aggregate is an "artificial unit," that is, one made up by

throwing together entirely unrelated groups. As an illustration of paradromy we take the array of marriage frequencies for the six cities, Barcelona, Birmingham, Boston, Leipzig, Melbourne and Rome, for the decade 1899-1908. By marriage frequence we mean the ratio of the number married (twice the number of marriages) to population.

For the six cities taken as a whole, with a total population of about three million, the marriage frequence $y_{o,k}$ varies between 18.00 and 19.02 per cent with an average of 18.38 per cent. The dispersion coefficient Q_o is 3.17. For the six cities taken singly in the above order, each with a population of about half a million, the values of Q_i are 2.69, 4.32, 4.17, 2.88, 3.76 and 2.72, with an average 3.42, somewhat higher than Q_o This result is a direct contradiction of the statement of Lexis that a narrowing field of observation reduces the value of Q. Lexis, without giving the matter much thought, worked with the hypothesis that isodromy, or at least a decided homodromy, always existed. In our example, however, we have paradromy, if not antidromy, for we find γ to be -0.054. Corresponding to this, we have β_o less than each of the values β_1 to β_6, for β_o approximates 0.0167 while β_i, $i = 1$ to 6, lies between 0.0334 and 0.0563. The quadratic mean of these quantities is 0.0450.

It is of prime interest to investigate for paradromy the theoretical relation of β_o to the quadratic mean of the values β_1, β_2 . . . β_n and of Q_o to the quadratic mean of Q_1, Q_2 . . . Q_n , for the case m_i = const. = m . In this case, $m_o = nm$, and if O is substituted for γ in (10) we have

$$\beta_o^2 = \frac{1}{n^2}\sum_{i=1}^{n}\beta_i^2 \quad \text{, whence} \quad \beta_o = \frac{1}{\sqrt{n}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\beta_i^2}$$

At the same time, we find on the one hand, from (9), the ex-

pected value

$$Q_0^2 = 1 + \frac{z}{z-1} m_0 \beta_0^2 ,$$

or

$$Q_0^2 = 1 + \frac{z}{z-1} \frac{m}{n} \sum_{i=1}^{n} \beta_i^2$$

and on the other hand

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1 + \frac{z}{z-1} \frac{m}{n} \sum_{i=1}^{n} \beta_i^2$$

whence

$$Q_0 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} Q_i^2}$$

In the marriage frequence example, where the quantities m_i, though not equal, differ very little from one another, we have the values already found

$$\beta_0 = 0.0167 \text{ and } \qquad Q = 3.17$$

to compare with the values

$$\frac{1}{\sqrt{n}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \beta_i^2} = 0.0184$$

and

$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} Q_i^2} = 3.49$$

The differences $0.0167 - 0.0184 = -0.0017$ and $3.17 - 3.49 = -0.32$ are explained partly by the fact that the assumption $m_i = \text{const.}$ is not exactly in accord with the facts, and partly because para-

dromy is really not present as assumed, but only a weak antidromy. This last should, however, be considered as due to chance. The artificial character of a total aggregate shows itself in paradromy.

Of the two quantities Q and β, only the latter can be considered as a proper measure of the stability of a statistical frequency—more exactly, of the corresponding fundamental probability. And, since on account of formulas (11) and (12), the total aggregate can never show a higher value of β than the average for the partial aggregates (because the upper limit for γ is 1), we obtain a glimpse of the question of the connection between stability and homogeneity.

The idea of homogeneity as we here understand it has reference to the result of the decomposition of a statistical aggregate according to some attribute or complex of attributes. The aggregate may consist of s elements, say s human beings and the decomposition may yield N sub-aggregates containing s', s'' . . . elements. Let some event A be observed x times in the total aggregate and x', x'', . . . times in the sub-aggregates. If we find the relative frequencies

$$y = \frac{x}{s}, \qquad y' = \frac{x'}{s'}, \qquad y'' = \frac{x''}{s''}, \ldots\ldots$$

then, on account of the two identities, $s' + s'' + \ldots = s$, and $x' + x'' + \ldots = x$, we have the relation

$$y = \frac{s'y' + s''y'' + \cdots\cdots}{s' + s'' + \cdots\cdot}$$

The "general frequency" then appears as the weighted arithmetic mean of the "special frequencies," y', y'', . . .

The theory of probabilities, with more or less assurance, furnishes us a criterion for deciding whether or not the deviations of the quantities y', y'', . . . from y are due to chance.

If they are not due to chance we say that the total aggregate "reacts" to the decomposition in question and that the attribute or complex of attributes which governs the decomposition is "relevant." If they are due to chance, we say that the total aggregate does not react to the decomposition and that the attribute is "indifferent."

According to the standpoint of the theory of probability, the relative frequencies y, y', y'' . . . as also the quotients $\frac{s'}{S}$, $\frac{s''}{S}$, . . . can be considered as approximations of distinct probabilities. If we designate the two series of probabilities thus inferred by p, p', p'', . . . and g', g'', . . . respectively, we find

$$(13) \qquad p = g'p' + g''p'' + \cdots$$

and the character of the attribute in question as relevant or indifferent finds expression in the fact that the "special probabilities" p', p'', . . . either differ from one another or are all equal to p, the "general probability."

For every ample enough complex of attributes we can imagine the decomposition going on and on by applying one attribute of the complex after another. Finally a point is reached where the sub-aggregates no longer react to further decomposition, or, expressed otherwise, the supply of relevant attributes is exhausted, and the probabilities p', p'', . . . which are associated with these sub-aggregates are called "elementary probabilities." In this case we say that the sub-aggregates themselves are "completely homogeneous" with reference to the event A.

The total aggregate—still in reference to A—is the more diversified the more the elementary probabilities p', p'', . . . differ among themselves, that is, the more they differ from p. It is reasonable to take as a measure of this diversity the expression

δ , defined by

$$\delta^2 = g'(p'-p)^2 + g''(p''-p)^2 + \cdots \tag{14}$$

Diversity and homogeneity are antithetical notions; the more undiversified the aggregate, the more it is homogeneous, and vice versa.

In order to apply this view of homogeneity, now considered for itself, to the procedure and the examples which we have brought forward in the discussion of stability, we must disregard the time fluctuations of the probabilities in question. That is, we do not use the quantities $p_{i,k}$ but fix attention on the probabilities p_i which refer to an individual time interval of n partial intervals—say a decade. By carrying out repeatedly the decomposition according to formula (13), the quantities p_i , p_o not included may be expressed in the form

$$p_i = g_i' p_i' + g_i'' p_i'' + \cdots$$

where p_i' , p_i'' . . . are elementary probabilities. Corresponding to formula (14), we have

$$\delta_i^2 = g_i'(p_i'-p_i)^2 + g_i''(p_i''-p_i)^2 + \cdots \tag{15}$$

If we designate the proportion of the i th partial aggregate to the total aggregate by c_i , that is, if we let $\frac{s_i}{s_o} = c_i$, we find

$$p_o = \sum_{i=1}^{n} c_i p_i$$

and at the same time

$$(16) \qquad \delta_0^2 = \sum_{i=1}^{n} \left\{ c_i g_i' (p_i' - p_0)^2 + c_i g_i'' (p_i'' - p_0)^2 + \cdots \right\}$$

The number of summands in (16) is nN, since there are n partial aggregates and each of these is a totality of N sub-aggregates. It may easily occur that some of the nN elementary probabilities are equal and this is expected in connection with elementary probabilities which are associated with similar sub-aggregates. But even in the most extreme case, where the elementary probabilities are equal without exception, we cannot say that the probabilities p_i are all alike. This can occur only when the values g_i', g_i'', . . . are independent of i. This highly improbable case is excluded from our discussion. We have then

$$(17) \qquad \sum_{i=1}^{n} c_i (p_i - p_0)^2 > 0$$

From (15) and (16), we have the following:

$$g_i' (p_i' - p_0)^2 + g_i'' (p_i'' - p_0)^2 + \cdots = \delta_i^2 + (p_i - p_0)^2$$

$$\delta_0^2 = \sum_{i=1}^{n} c_i \delta_i^2 + \sum_{i=1}^{n} c_i (p_i - p_0)^2$$

so that, on account of (17)

$$\delta_0^2 > \sum_{i=1}^{n} c_i \, \delta_i^2$$

and *a fortiori*

$$(18) \qquad \delta_0 > \sum_{i=1}^{n} c_i \, \delta_i$$

The total aggregate is then under all circumstances less homogeneous than the partial aggregates are on the average

This statement might possibly correspond to the every-day meaning of the word "homogeneity," which carries with it no precise quantitative idea. Indeed, when we consider that in the case of the total aggregate we have to take into account not only the lack of homogeneity within the partial aggregates, but also the diversity with which the partial aggregates may make up the whole, we are inclined to say that the total aggregate is less homogeneous than any of its parts. With that idea, however, we do not hit upon the right thing as far as our mathematical criterion of homogeneity is concerned. The inequality (18) says only, that the average of the values $\delta_1, \delta_2, \ldots \delta_n$ is less than δ_0, not that each one is less than δ_0.

In our foregoing discussion of stability as measured by the relative essential fluctuation component, we found that for the total aggregate the stability was higher than the average for the partial aggregates, except for the case of isodromy, which in practice rarely occurs. Hence, there exists between homogeneity and stability an antagonistic relation—small homogeneity goes hand in hand with great stability. For example, the provinces into which a country may be divided will show, on the average, a greater homogeneity and at the same time a lesser stability in reference to an event A than will the country taken as a whole.

Again, the districts into which the provinces may be divided will on the average show a greater homogeneity associated with a still smaller stability. We can say that in general the homogeneity increases with the narrowing of the field of observation, while the stability decreases.

Is this to be considered as a warning against the all too popular diversification of statistical material which is being more and more accepted in research methods? Not in the least. That would be an obsolete point of view, as if the problem of statistics consisted in a search for most stable values. Rather does the opposition between homogeneity and stability give direction to business practice, especially to that branch of business which is in such close touch with statistics, namely insurance, where stability is of prime importance. It has been known for a long time that it contributes to the even tenor of the business side if the risks are as heterogeneous as possible. It is of advantage if the insured persons or things are spread relatively widely according to geographical and other points of view, instead of concentrating on a limited territory or few kinds of risks.

Accordingly, even if this thesis, that an antagonistic relation exists between homogeneity and stability, seems surprising and strange, we find on closer consideration that the theory agrees with a practice which has instinctively grasped the true situation. It is now twelve years since I had the first opportunity to explain at greater length than here the foregoing developed ideas and with the verifying data to present them to my colleagues.[1] As far as I know, only one of these has taken a definite stand in the matter. This is John Maynard Keynes.[2] He makes the charge against me, that instead of clearing up a very simple matter, I have befogged it with a profusion of mathematical formulas

---

[1]Homogeneitat und Stabilitat in der Statistik, in the Skandinavisk Aktuarietidskrift, 1918, pages 1-81, Upsala.

[2]A treatise on probability, London, 1921, pages 403-405.

and new technical terms, and he believed that he could show this best by an example of my own from the field of insurance. In referring to this example, Keynes thought that the distinction made by myself in a much earlier publication between a general probability p and the special probabilities $p_1, p_2, \ldots$ was the one in question, where

$$p = \frac{z_1}{z} p_1 + \frac{z_2}{z} p_2 + \cdots$$

Keynes further expressed himself as follows:

> "If we are basing our calculations on p and do not know p_1, p_2, etc., then these calculations are more likely to be borne out by the result if the instances are selected by a method which spreads them over all the groups 1, 2, etc., than if they are selected by a method which concentrates them on group 1. In other words the actuary does not like an undue proportion of his cases to be drawn from a group which may be subject to a common relevant influence *for which he has not allowed.* If the *a priori* calculations are based on the average over a field which is not homogeneous in all its parts, greater stability of result will be obtained if the instances are drawn from all parts of the non-homogeneous total field, than if they are drawn now from one homogeneous subfield and now from another. This is not at all paradoxical. Yet I believe, though with hesitation, that this is all that Von Bortkiewicz's elaborately supported mathematical conclusion amounts to."

Suppose, for example, that a fire insurance company insures

[1]Here z refers to a series of "equally likely events," which is broken up into groups of $z_1, z_2, \ldots$ equally likely events. Hence $z = z_1 + z_2 + \cdots$

two kinds of buildings, dwellings and factories, which are classified as different grades of fire risks, for insurance premiums which are not graded. The premium is to be calculated per unit on the supposition that the risks in the two categories are divided in a definite proportion. Then, according to Keynes, a greater stability in the business is guaranteed if every year dwellings as well as factories are insured, than if in one year only dwellings and in another year only factories are insured. This is certainly true and requires no lengthy argument. But it has nothing whatever to do with my thesis of the antagonistic relation between stability and homogeneity.

To give an example which does illustrate my theory, think of three insurance companies, A, B, and C. A insures only dwelling houses, B only factories, while C insures both. The premiums in A, B. and C are different because of the different classes of risks. It is assumed in C that there is no grading of premiums. A premium per unit is charged which is calculated according to the relative number of the two risks. The premium is to be just high enough so that for a period of years, allowing for variations due to chance, the damages are just covered. In the course of this period, the danger of fire varies from year to year, showing gains in some years, losses in others. Such fluctuations of fire hazard would correspond in my scheme to the variations of the probabilities $p_{i,k}$ with respect to k, while $p_{1,k}$ is associated with A, $p_{2,k}$ with B, and $p_{0,k}$ with C. And in accord with my theory that, except in the case of isodromy, the values $p_{0,k}$, relatively speaking, show weaker variations than $p_{1,k}$ and $p_{2,k}$ do on the average, the insurance company C would show relatively smaller fluctuations of fire damage from one year to another, resulting in a more stable business than would be shown by the average of A and B. The mixed character of the risks would be conducive to greater stability. In the case of C a certain compensation of effects would take place

which the time variations of the two-sided fundamental probabilities would make manifest on the business side.[1] But Keynes says nothing of these variations. He simply missed the point of my argument and his remarks were not relevant.

It is to be hoped that the new exposition of my theory, although, or because, it is essentially shorter than the older one, will give no cause for a similar misunderstanding.

---

[1]This compensation would also appear in the more complicated case where the proportions of the risks in C are not unchangeable as is assumed in the text, but would change from year to year (the premium being adjusted accordingly). We need not go further into this matter because, in my theory, the composition of $s_{\cdot,k}$ out of the component parts $s_{i,k}$ is considered as fixed. In my examples, this composition varied, but the fluctuations were insignificant in comparison to the variations of the values $p_{i,k}$. See Skandinavisk Aktuarietidscrift, pages 69-70.

L. v. Bortkiewicz.

BAYES' THEOREM

An Expository Presentation*

By

EDWARD C. MOLINA

American Telephone and Telegraph Company

Bayes' theorem made its appearance as the ninth proposition in an essay which occupies pages 370 to 418 of the Philosophical Transactions, Vol. 53, for 1763. An introductory letter written by Richard Price, "Theologian, Statistician, Actuary and Political Writer,"[1] begins thus:

> "I now send you an essay which I have found amongst the papers of our deceased friend, Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved."

A few lines further on Price says:

> "In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know

*Read before the American Statistical Association during the meeting of the American Association for the Advancement of Science in Cleveland, Ohio, December, 1930.

[1] These titles are associated with the name of Price in the frontispiece portrait of him bound with the December, 1928, issue of *Biometrika*.

nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times."

. . . .

"Every judicious person will be sensible now that the problem mentioned is by no means merely a curious speculation in the doctrine of chances, but necessary to be solved in order to assure a foundation for all our reasonings concerning past facts, and what is likely to be hereafter."

No one will dispute the importance ascribed to Bayes' problem by Price; in fact, a paper by Karl Pearson on an extension of Bayes' problem is entitled "The Fundamental Problem of Practical Statistics." Opinions differ, however, as to the validity and significance of the solution submitted in the essay for the problem in question. In view of this situation I shall limit myself today to an exposition of the fundamental characteristics of the problem Bayes' theorem deals with and shall give no consideration to its interesting applications.

The exposition may be outlined as follows: after specifying the class of problems to which Bayes' theorem pertains, I shall:

I. Discuss briefly two problems, each of which will emphasize one of two kinds of *a priori* probabilities which should be constantly borne in mind when Bayes' theorem is under consideration,

II. Partially analyze a certain ball-drawing problem which will not only serve as an introduction to the algebra of Bayes' theorem but will later help to throw light on its significance,

III. Present Bayes' problem and the related theorem.

IV. Make some remarks on the value of the theorem and the controversies which it raised.

In carrying out this plan I shall find it convenient to ignore the historic order of events.

When probability is the subject under consideration one an-

ticipates problems such as: A coin is about to be tossed 15 times; What is the probability that heads will turn up seven times? A sample of 100 screwdrivers is to be taken from a case containing 1000 screwdrivers of which 300 are known to be defective; what is the probability that the sample will contain 25 defectives?

These are direct, or *a priori, probability problems*. In each of them the nature of a game, or an experiment, is specified in advance and then a question is asked relating to one, or more, of the possible outcomes of the game or experiment. Problems of this type have occupied the attention of mathematicians since the days of Pascal and Fermat, the creators of the mathematical theory of probability.

An inverse class of problems of great practical significance, called *a posteriori* probability problems, came into prominence with the publication of Bayes' essay. In these we find specified the result or outcome of a game which has been played, whereas the question then asked is whether the game actually played was one or some other of several possible games. This type of problem is usually stated as follows:

> "An event has happened which must have arisen from some one of a given number of causes; required the probability of the existence of each of the causes."

I

Consider this example: During his sophomore year Tom Smith played on both the baseball and football varsity teams; we have been informed that he broke his ankle in one of the games; what are the *a posteriori* probabilities in favor of baseball and football, respectively, as the baneful cause of the accident?

Evidently the answer depends on the number of baseball and football games played during their respective seasons and also on the likelihood of a man breaking an ankle in one or the other of

these two games. As a concrete case assume that:

1. At Smith's college an equal number of baseball and football games are played per season;
2. Statistical records indicate that if a student participates in a baseball game the probability is 2/100 that he will break an ankle and that, likewise, the probability is 7/100 for the same contingency in a football game.

In view of the first of these two assumptions our conclusions as to the cause of the accident may be based entirely on the information contained in the second assumption. The odds are two to seven, so that the *a posteriori* probabilities regarding the two admissible causes are:

$$\text{For baseball, } 2/(2+7) = 2/9.$$
$$\text{For football, } 7/(2+7) = 7/9.$$

Now consider this other example. A lone diner amused himself between courses by spinning a coin. We elicited from the waiter that in 15 spins heads turned up seven times. Moreover, from our point of observation, the size of the coin indicated that it was either a silver quarter or a ten-dollar gold piece. What are the *a posteriori* probabilities in favor of the silver quarter and the gold piece, respectively?

If the lone diner were a professor from one of our eastern universities we would not hesitate a moment in declaring that the coin spun was a quarter. But it happens that the gentleman was a member of the Cleveland Chamber of Commerce, dining at the Bankers' Club. We must, therefore, give the matter more careful consideration. The number of quarters and gold pieces usually carried by a banker and the probabilities of obtaining the observed result by spinning coins are relevant; let us assume, therefore, that:

1. The small change purse of a Cleveland financier contains, on the average, ten-dollar gold pieces and quarters in the ratio of

eight to three.

Moreover, we may assume (in fact we know) that:

2. If either a quarter or a gold piece is spun 15 times, the probability that heads will turn up seven times is approximately 1/5.

The second of these two items of information makes the *a posteriori* probabilities depend entirely on the first item. Clearly the odds are eight to three and we conclude:

For a quarter, *a posteriori* probability $=3/(3+8) = 3/11$.
For a goldpiece, *a posteriori* probability $=8/(3+8) = 8/11$.

Now regarding the general *a posteriori* problem,

> "An event has happened which must have arisen from some one of a number of causes; required the probability of the existence of each of the causes,"

what do the two examples we have just considered suggest? In both problems we inquired into:

1. The frequency with which each of the possible causes is met BEFORE THE OBSERVED EVENT HAPPENED. This frequency is called the *a priori existence* probability for the corresponding cause.
2. The probability that a cause, if brought into play, would reproduce the observed event. This probability will hereafter be referred to as the *a priori productive* probability for the cause in question.

In the case of the broken ankle, the *a priori existence* probabilities were equal and took no part in our conclusion; we based the *a posteriori* probabilities entirely on the *a priori productive* probabilities. We did just the opposite with reference to the coin spun by the Cleveland financier; on account of the equality of the *a priori productive* probabilities we deduced *a posteriori* prob-

abilities in terms of the unequal *a priori* existence probabilities.

It is apparent that our two examples represent extreme cases. In general, the solution of an inverse or *a posteriori* problem, involving a number of causes, one of which must have brought about a certain observed event, depends on both sets of direct, or *a priori* probabilities. Those of the first set give the frequency with which the various causes were to be expected before the observed result occurred; those of the second set give the frequencies with which the observed result would follow from the various causes if each were brought into play.

II

Bearing in mind the two distinctly different sets of *a priori* probabilities required in arriving at *a posteriori* conclusions regarding the possible causes of an observed event, we must now give some thought to the algebra of the subject before taking up Bayes' problem and theorem. For this purpose consider the following bag problem:

A bag contained M balls, of which an unknown number were white. From this bag N balls were drawn and of these T turned out to be white. What light does this outcome of the drawings throw on the unknown ratio of the number of white balls to the total number of balls, M, in the bag? Let x be this unknown ratio.

Two cases of this problem may be considered:

Case 1.—After a ball was drawn it was replaced and the bag was shaken thoroughly before the next drawing was made.

Case 2.—A drawn ball was not replaced before the next drawing.

These two cases become essentially identical when the total number of balls in the bag is very large compared with the number drawn. Case 1 will serve as an introduction to Bayes' prob-

lem; later we will find it highly desirable to consider Case 2.

We are confronted with $(M+1)$ possible hypotheses or causes before the drawings took place:

1 – the unknown value of x is $x_0 = 0/M$,
2 – the unknown value of x is $x_1 = 1/M$,
3 – the unknown value of x is $x_2 = 2/M$,
. . . . . .
$k+1$ – the unknown value of x is $x_k = k/M$,
. . . . . .
$M+1$ – the unknown value of x is $x_M = M/M = 1$.

Let $w(x_k)$ be the *a priori* existence probability for the k'th hypothesis; by this is meant the probability in favor of the k'th hypothesis based on whatever information was available regarding the contents of the bag prior to the execution of the drawings.

Let $B(T, N, x_k)$ be the *a priori productive* probability for the k'th hypothesis; by this is meant the probability of obtaining the observed result (T whites in N drawings) when the value of x is k/M.

Then, the *a posteriori* probability, or probability after the observed event, in favor of the k'th hypothesis is

$$(1)^1 \qquad p_k = \frac{w(x_k)\, B(T, N, x_k)}{\sum_{k=0}^{M} w(x_k)\, B(T, N, x_k)}$$

For Case 1 of our bag problem we have

$$B(T, N, x_k) = \binom{N}{T} x_k^T (1-x_k)^{N-T}$$

[1] This is the Laplacian generalization of Bayes' formula, although in some textbooks it is referred to as "Bayes' Theorem." A relatively short demonstration of it is given by Poincaré in his *Calcul des Probabilités*. See also Fry, *Probability and its Engineering Uses*, Art. 49.

where $\binom{N}{T}$ represents the number of combinations of N things taken T at a time. Substituting in (1), we obtain, after canceling from numerator and denominator the common factor $\binom{N}{T}$,

$$(2) \qquad p_k = \frac{w(x_k)\, x_k^{T}\, (1-x_k)^{N-T}}{\sum_{k=0}^{M} w(x_k)\, x_k^{T}\, (1-x_k)^{N-T}}$$

If in equation (2) we give k successively the values a, $a+1$, $a+2$, . . . $b-1$, b and add the results, we have

$$p_a + p_{a+1} + \cdots\cdots + p_b$$

or

$$(3) \qquad p(x_a x_b) = \frac{\sum_{k=a}^{k=b} w(x_k)\, x_k^{T}\, (1-x_k)^{N-T}}{\sum_{k=0}^{M} w(x_k)\, x_k^{T}\, (1-x_k)^{N-T}}$$

for the *a posteriori* probability that the unknown ratio of white to total balls in the bag lies between a/M and b/M, both inclusive.

III

BAYES' PROBLEM

Consider the table represented by the rectangle $ABCD$ in Fig. 1. On this table a line OS was drawn parallel to, but at an unknown distance from, the edges AD and BC. Then a ball was rolled on the table N times in succession from the

edge AD toward the edge BC. As indicated in the figure it was noted that T times the ball stopped rolling to the right of the line OS and $N-T$ times to the left of that line.

What light does this information shed on the unknown distance from AD to OS? In more technical terms, what is the *a posteriori* probability that the unknown position of the line OS lies between any two positions in which we may be interested?

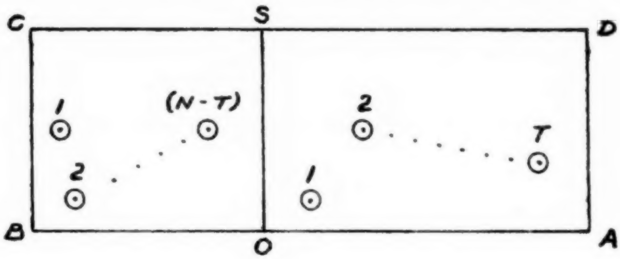


Fig. 1.

Each rolling of the ball was executed in such a manner that the probability of the ball coming to rest to the right of OS is given by the unknown ratio of the distance OA to the length BA of the table; likewise, the probability of the ball stopping to the left of OS is given by the ratio of the distance BO to the length BA.

$$\text{Set } x = OA/BA, \quad 1 - x = BO/BA.$$

The only difference between this problem and the bag of balls problem is that now the possible values of x are not restricted to the finite set $0/M, 1/M, 2/M, \ldots (M-1)/M, M/M$; in the table problem x may have had any value whatever between the limits of 0 and 1. Therefore equation (3) will answer the question asked provided we substitute definite integrals in place of the finite summations. This substitution gives us, for the de-

sired *a posteriori* probability that x had a value between x_1 and x_2, the formula

$$(4) \qquad p(x_1, x_2) = \frac{\int_{x_1}^{x_2} w(x)\, x^T (1-x)^{N-T} dx}{\int_0^1 w(x)\, x^T (1-x)^{N-T} dx}$$

Equation (4) is useless until the form of the *a priori existence* function $w(x)$ is specified; this depends on the *way* in which the line OS was drawn. Bayes assumed that the line OS, of unknown distance from AD, was drawn through the point of rest corresponding to a preliminary roll of the ball. This amounts to postulating that all values of x, between 0 and 1 were *a priori* equally likely. In other words, with Bayes, the *a priori existence function* $w(x)$ was a constant which, therefore, did not have to be taken into consideration.[1] Thus, instead of equation (4), Bayes gave the equivalent of the following restricted formula:

$$(5) \qquad p(x_1, x_2) = \frac{\int_{x_1}^{x_2} x^T (1-x)^{N-T} dx}{\int_0^1 x^T (1-x)^{N-T} dx}$$

I say "the equivalent of" (5) because in Bayes' day definite integrals were expressed in terms of corresponding areas.

Equation (5) constitutes Proposition 9 of the essay, but is usually referred to as Bayes' theorem.

---

[1] The existence function $w(x)$ does not appear either explicitly or implicitly anywhere in Bayes' essay. This fact raises the question as to whether or not Bayes had any notion of the *general* problem of causes.

IV.

Equation (5) is a very beautiful formula; but we must be cautious. More than one high authority has insinuated that its beauty is only skin deep. Speaking of Laplace's generalization and extension of the theorem, George Chrystal, the English mathematician and actuary, closed a severe attack on the whole theory of *a posteriori* probability[1] with the statement that "Practical people like the Actuaries, however much they may justly respect Laplace, should not air his weaknesses in their annual examinations. The indiscretions of great men should be quietly allowed to be forgotten."

Chrystal's advice as to the attitude one should assume toward "the indiscretions of great men" is excellent, but in the case under consideration, it was the plaintiff rather than the defendant who committed indiscretions; this is discussed in a paper by E. T. Whittaker[2] entitled "On Some Disputed Questions of Probability."

The discussions and disputes, which began shortly after the birth of the formula in 1763 and which have not as yet subsided, may be divided into two classes:

1. Discussions concerning problems in which it is known that the *a priori* existence function is not a constant.
2. Discussions concerning problems in which nothing whatever is known concerning the *a priori* existence function.

The discussions of Class 1 are out of order in so far as Bayes' theorem is concerned; recourse should be had to formula (4), Laplace's generalization of the Bayes' theorem, when it is known that $\omega(x)$ is not a constant. Failure to differentiate

[1] "On Some Fundamental Principles in the Theory of Probability," *Transactions of the Actuarial Society of Edinburgh*, Vol. 11, No. 13.

[2] *Transactions of the Faculty of Actuaries in Scotland*, Vol. VIII, Session 1919-1920.

explicitly between equations (4) and (5) has created a great deal of confusion of thought concerning the probability of causes. The discussions of Class 2 have centered on what Boole called "the equal distribution of our knowledge, or rather of our ignorance," that is to say "the assigning to different states of things of which we know nothing, and upon the very ground that we know nothing, equal degrees of probability." Regarding the legitimacy of this procedure Bayes himself contributed a very important scholium, which appeared in his essay on pages 392 and 393. The argument in this scholium, based on a corollary to Proposition 8 of the essay, may be summarized as follows:

Assuming that all values of x are *a priori* equally likely and that the N throws of a ball on the table have *not yet* been made, the probability that T times the ball will rest to the right of OS and that the remaining $N-T$ times it will rest to the left of OS is (as shown in the corollary)

$$(6) \qquad p = \int_0^1 \binom{N}{T} x^T (1-x)^{N-T} dx = \frac{1}{N+1}$$

a result in which T does not appear. In other words, any assigned outcome for the throws is no more, or no less, likely than any other outcome, if *a priori* all values of x are equally likely. But, wrote Bayes in the scholium, when we say that we have no knowledge whatever *a priori* regarding the ratio x, do we not really mean that we are in the dark as to what will be the outcome when we proceed to make N throws? If so, then equation (6) justifies the assumption that *a priori* all values of x are equally likely.

To clinch his argument it must be shown that the converse of equation (6) is true. That is, it must be shown that, if any outcome of throws *not yet* made is as likely as any other, then

any value of x is *a priori* as likely as any other. This converse theorem was submitted to Dr. F. H. Murray, who obtained an elegant proof based on a theorem of Stieltjes.[1]

In view of Bayes' corollary and his scholium, an analysis of our bag problem with reference to the "equal distribution of our knowledge, or ignorance" is in order.

Consider again Case 1 where each drawn ball is replaced in the bag before the next drawing is made.

Assuming each of the $(M+1)$ permissible hypotheses to be *a priori* equally likely, the probability that N drawings, *not yet* made, will result in T white and $N-T$ black balls is

$$P = \sum_{k=0}^{M} \frac{1}{M+1} \binom{N}{T} \left(\frac{k}{M}\right)^{T} \left(1-\frac{k}{M}\right)^{N-T} \tag{7}$$

Equation (7) is not, in general, independent of T[2] so that any one assigned outcome of N drawings is not as likely as any other outcome. This result is disturbing; at first sight it seems to discredit Bayes' scholium. We must, therefore, look into the the matter more closely.

Bayes' problem corresponds to drawings from a bag containing an infinite number of balls. Therefore, even if drawn balls are replaced, the chance of a particular ball being drawn more than once is zero. But when N drawings with replacements are made from a bag containing a *finite* number, M, of balls, we are by no means certain of drawing N different balls;

---

[1] *Bulletin of the American Mathematical Society*, February, 1930.

[2] Consider, for example, the case of $M = 2$. Equation (7) reduces to

$$P = \frac{1}{3}\left(\frac{1}{2}\right)^{N}\binom{N}{T}$$

a result which is not independent of T.

a particular white ball may be drawn several times over, and, likewise, a particular black ball may appear more than once. It is not surprising, therefore, that Case 1 of the bag problem does not confirm Bayes' corollary.

Consider now Case 2, where the drawn balls are not returned to the bag. If k of the total balls are white and the rest black, the probability that a sample of N balls from the bag will contain T white and $N-T$ black is

$$\binom{k}{T}\binom{M-k}{N-T}\Big/\binom{M}{N}$$

Hence, if the permissible values 0, 1, 2, 3, . . . M for k are all equally likely *a priori*, we obtain instead of (7),

$$P=\sum_{k=0}^{M}\left(\frac{1}{M+1}\right)\binom{k}{T}\binom{M-k}{N-T}\Big/\binom{M}{N}=\frac{1}{N+1} \tag{8}$$

a result independent of any assigned value for T and identical with the result in the corollary to Proposition 8 of the essay.

SUMMARY

Bayes' theorem is the answer to a special case of the general problem of causes. The special case postulates that the *a priori* existence probabilities for the various admissible causes of an observed event are equal.

In the essay Bayes recommends that his theorem be adopted whenever we find ourselves confronted with total ignorance as to which one of several possible causes produced an observed event. To justify this recommendation Bayes takes the attitude that: A state of total ignorance regarding the causes of an ob-

served event is equivalent to the same state of total ignorance as to what the result will be if the trial or experiment has not yet been made. This interpretation is a generalization of the fact that in his billiard table problem, the assumption of equal likelihood for all possible positions of the line OS, gives equal probabilities for the various possible outcomes of a set of N ball rollings not yet made.

Laplace, Poincaré and Edgeworth[1] have shown that the *a priori existence* function $w(x)$, which appears in the Laplacian generalization of Bayes' theorem, is of negligible importance when the numbers N and T are large. Therefore, when this condition holds, one need not hesitate to use Bayes' restricted formula for the solution of a problem of causes.

The transmission, by Price, of Bayes' posthumous essay to the Royal Society marked an epoch in the history of the literature on probability theory. As mentioned at the beginning of this paper, Karl Pearson has called the extension of Bayes' problem the "Fundamental Problem of Practical Statistics."

[1] Laplace: "Oeuvres," Vol. 9, p. 470. Poincaré: "Calcul des Probabilités," 2d edition, p. 255. Bowley: "F. Y. Edgeworth's Contribution to Mathematical Statistics," pp. 11 and 12.

E. C. Molina

ON CERTAIN PROPERTIES OF FREQUENCY DISTRIBUTIONS OBTAINED BY A LINEAR FRACTIONAL TRANSFORMATION OF THE VARIATES OF A GIVEN DISTRIBUTION

By

H. L. Rietz

Considerable evidence has been presented by R. A. Fisher[1] to show that, by an appropriate transformation $z = f(r)$ of small sample correlation coefficients $r(-1 \leqq r \leqq 1)$ distributed in accord with a decidedly skew frequency curve, values of z are obtained which are distributed nearly in a normal distribution. In fact, the approach of the distribution of z to normality seems sufficiently rapid to justify the use of the probable error of z in many applications as if it were normally distributed. Such a change in the character of the distribution of an important statistic suggests the further study of properties of the distribution of variables obtained by applying rather simple transformations to variates distributed from −1 to +1 in accord with a given frequency function. In a previous paper,[2] the writer has dealt with a similar problem when each variate of a given unimodal distribution of any finite range is replaced by a given power of the variate.

Consider a positive unimodal continuous frequency function

[1] Metron, Vol. 1, Part 4 (1921) pp. 3-32.

[2] Proceedings of the National Academy, Vol. 13, No. 12 (1927), pp. 817-820.

$y = \psi(x)$ of a system of variates $x_1, x_2, \ldots x_n$ with a range of -1 to $+1$, with $\psi(-1)=\psi(1)=0$, with a single mode at some point, say at $x=b$ $(-1 < b < 1)$, and with the derivative $\psi'(x)$ continuous. More precisely, we assume that $\psi(x)$ is positive except at the end points of the interval -1 to $+1$, where it is zero, and that $\psi'(x)$ changes from positive to negative at $x=b$, and is non-negative or non-positive at any point $x=a$ according as a is less or greater than b.

It is the main object of the present paper to consider certain properties of the distribution of variates $u_i = (ex_i + f)/(gx_i+h)$ obtained by a linear fractional transformation of the x's, where e, f, g, and h are real numbers so selected that $u = (ex+f)/(gx+h)$ is continuous from $x=-1$ to $x=1$.

When $g=0$, we have the case of the linear transformation which simply has an effect equivalent to a change of origin and of unit of measurement. As we are not in the present problem much interested in such a simple transformation, we shall, in general, assume $g \neq 0$. Moreover, we take g positive, since this involves no loss of generality.

We shall, except as otherwise stated, restrict our considerations to the interval for u that corresponds to $-1 \leq x \leq 1$, and to such transformations that the derivative of u with respect to x is finite for each value of x and that u increases when x increases. These restrictions require that

$$\frac{du}{dx} = \frac{he-fg}{(gx+h)^2}$$

where $g < |h|$ and where the determinant

$$he-fg = \begin{vmatrix} e & f \\ g & h \end{vmatrix} > 0 \tag{1}$$

Starting then with

$$u = \frac{ex+f}{gx+h}, \tag{2}$$

we have

$$x = \frac{f-hu}{gu-e}. \tag{3}$$

Next, let

$$v = \phi(u) \tag{4}$$

be the frequency function of the new variates u. Then we may write[1]

$$v = \phi(u) = \psi\left(\frac{f-hu}{gu-e}\right)\cdot\frac{he-fg}{(gu-e)^2}. \tag{5}$$

Since $he-fg > 0$, we know that v is positive throughout the interval in which we are interested except that $V = 0$ at the end points. From (5) it seems that the new distribution function may possibly become infinite when $u = e/g$, but the question then arises as to whether e/g is an admissible value of u.

We shall prove that e/g is not an admissible value of u by showing that u cannot take the value e/g within the interval $u = (f-e)/(h-g)$ to $u = (e+f)/(g+h)$ wherein u lies when $-1 \leqq x \leqq 1$. In this connection we shall also establish some inequalities that will be found useful in the consideration of certain properties of the new distribution. Consider first the cases in which $g+h$ is positive.

Then since $eh > fg$, we have $eh+eg > fg+eg$.

Divide by $g(g+h)$, and we have $\frac{e}{g} > \frac{f+e}{g+h}$ Hence,

[1] cf. Annals of Mathematics, vol. 23, No. 4 (1922), pp. 293-4.

e/g is too large when $g+h$ is positive to be an admissible value of u.

Consider next the cases in which $g+h$ is negative. In this case, $h<0$ since $g>0$. Hence $g-h>0$. Then since $eh>fg$, we have $eh-eg>fg-eg$. Divide by the positive number $g(g-h)$. This gives $\frac{-e}{g}>\frac{f-e}{g-h}$ and $\frac{e}{g}<\frac{e-f}{g-h}$.
Hence, when $g(g+h)<0$, e/g is too small to be an admissible value of u.

To summarize with $g>0$, we have shown that:

(a) When $g+h$ is positive, e/g is too large to be an admissible value of u.

(b) When $g+h$ is negative, e/g is too small to be an admissible value of u.

Returning now to the consideration of our frequency function $v=\psi\left(\frac{f-hu}{gu-e}\right)\cdot\frac{he-fg}{(gu-e)^2}$ in (5), we obtain

$$(6)\quad \frac{dv}{du}=\frac{(he-fg)^2}{(gu-e)^4}\psi'\left(\frac{f-hu}{gu-e}\right)-\frac{2g(he-fg)}{(gu-e)^3}\psi\left(\frac{f-hu}{gu-e}\right).$$

When u takes the value $(eb+f)/(gb+h)$ into which variates at the mode $x=b$ are transformed, we know that $\psi'\left(\frac{f-hu}{gu-e}\right)=\psi'(b)=0$.

By making use of the fact that $he-fg>0$, and the propositions (a) and (b) relating to the inadmissibility of e/g as a value of u in an examination of the right hand member of (6) for $u=(eb+f)/(gb+h)$, we establish the following proposition in regard to the sign of the derivative dv/du for the value of u which corresponds to the modal value of x

When $g+h\neq 0$, dv/du is positive or negative at $u=(eb+f)/(gb+h)$ according as $g+h$ is positive or negative.

The truth of this proposition follows readily by applying (a) and (b) to (6), remembering that g is positive and that ψ' (b) vanishes.

We shall show next in case $g+h > 0$, that dv/du is non-negative for all admissible values of u less than $(eb+f)/(gb+h)$. To see this from (6), note first that $\psi'[(f-hu)/(gu-e)]$ remains non-negative for $(f-hu)/(gu-e) < b$ or for u less than $(eb+f)/(gb+h)$, and note second that $g/(gu-e)^3$ is negative since e/g is too large to be an admissible value of u under the condition $g+h > 0$.

Next, in case $g+h < 0$, dv/du is non-positive for all values of $u > (eb+f)/(gb+h)$. To see this from (6), note first that $\psi'[(f-hu)/(gu-e)]$ remains non-positive for $(f-hu)/(gu-e) > b$ or for $u > \frac{(eb+f)}{(gb+h)}$, and note second that $g/(gu-e)^3$ is positive when $g+h<0$ because in this case $u > e/g$.

To summarize, when $g+h \neq 0$, we state the

> Theorem I. *When the derivative* dv/du *is positive for the value of* u *into which variates at the modal value* $x=b$ *transform, then* dv/du *is non-negative for all smaller values of* u. *Similarly, when* dv/du *is negative for the value of* u *into which variates at the modal value* $x=b$ *transform, then* dv/du *is non-positive for all larger values of* u.

Finally, we wish to inquire about a modal value for the frequency function $v=\phi(u)$ in (5). To this end, consider first the case in which dv/du is positive at $u=(eb+f)/(gb+h)$. At a point between $u=(eb+f)/(gb+h)$ and the upper bound of u, that is $(e+f)/(g+h)$, a maximum value of v occurs. To

see this, note when $u=(e+f)/(g+h)$ that $dv/du=\psi'(1)(g+h)^4/(he-fg)^2$ which is negative, or zero since $\psi'(1)$ is negative or zero. If it is negative, there is a maximum where the sign of the continuous first derivative changes from positive to negative. If dv/du is zero at $u=(e+f)/(g+h)$, it follows also that there is at least one maximum of $v=\phi(u)$ between $u=(eb+f)/(gb+h)$ and $u=(e+f)/(g+h)$ since $v=0$ at $u=(e+f)/(g+h)$ and v must have changed from an increasing positive function at $u=(eb+f)/(gb+h)$ to a decreasing function before becoming zero at $u=(e+f)/(g+h)$. Similarly, it may be shown that there is a mode at a value of $u<(eb+f)/(gb+h)$ whenever dv/du is negative at $u=(eb+f)/(gb+h)$.

We may then state the following:

Theorem II. *Given a unimodal continuous positive function* $y=\psi(x)$ *of variates* x, *with a range from* -1 *to* $+1$, *with a mode at* $x=b\,(-1<b<1)$, *with* $\psi(-1)=\psi(1)=0$, *and with the derivative* $\psi'x$ *continuous from* $x=-1$ *to* $x=1$, *then the frequency distribution* $v=\phi(u)$ *of variates* $u=(ex+f)/(gx+h)$ $(g>0)$ *has a mode at a value of* $u>(eb+f)/(gb+h)$ *when* $g+h>0$. *It has a mode at a value of* $u<(eb+f)/(gb+h)$ *when* $g+h<0$.

Since we have so restricted our transformation $u=\dfrac{(ex+f)}{(gx+h)}$ that the order of corresponding values is preserved, the transformation carries the median of the distribution of x's into the median of the distribution of u's, and we may state the following:

Corollary. *If* $y=\psi(x)$ *has its median and mode coincident at* $x=b$, *the frequency distribution* $v=\phi(u)$ *of* $u=(ex+f)/(gx+h)$ *has a modal value greater or less than its median according as* $g+h$ *is greater or less than zero.*

Thus far we have imposed the condition $g < |h|$. Let us next consider the cases in which $h = -g$ and $h = g$ instead of requiring that $g < |h|$. Consider first the case $h = -g$. In this case

(7) $$u = \frac{1}{g} \cdot \frac{ex+f}{x-1}$$

and

(8) $$\frac{du}{dx} = \frac{he-fg}{(gx+h)^2} = -\frac{e+f}{g(x-1)^2}.$$

Both u and du/dx become infinite as x approaches 1. Suppose e and f so chosen that u is an increasing function of x for the interval $-1 \leqq x < 1$, then u in (7) is an increasing function of x for the larger interval $-\infty < x < 1$; and it follows, for the case $h = -g$, that e/g is too small to be an admissible value of u when $-1 \leqq x < 1$, since it is the value of u when $x = -\infty$.

For the case $h = g$, we have

(9) $$u = \frac{ex+f}{g(x+1)}$$

and

(10) $$\frac{du}{dx} = \frac{e-f}{g(x+1)^2}.$$

Since u in (9) is an increasing continuous function of x for the interval $-1 < x < \infty$ wherever e and f are so selected that it is increasing for the sub-interval $-1 < x \leqq 1$, it follows, for $h = g$, that e/g, the value of u when $x = \infty$, is too large to be an admissible value of u when $-1 < x \leqq 1$. By making use of the fact that e/g is too small or too large

to be an admissible value of u according as $h=-g$ or $+g$, we readily obtain the following results from an examination of (6): The derivative dv/du given in (6) is positive at the point $u=(eb+f)/(gb+h)$ when $h=g$, and it is negative at this point when $h=-g$.

Moreover it readily follows as in the case where $g<|h|$ that when the derivative dv/du is positive for the value of u into which the modal $x=b$ transforms, then dv/du is non-negative for all smaller values of u, and when dv/du is negative for the value of u into which the modal value $x=b$ transforms, it is non-positive for all larger values of u.

Next, for the case $h=g$, a mode occurs for a value of $u>(eb+f)/(gb+h)$. This may be seen by noting that as x approaches 1 and as u takes corresponding values dv/du in (6) approaches the value $16\,g^2\,\psi'(1)/(e-f)^2$ which is negative or zero. The analysis given above for the corresponding case $g<|h|$ may be applied, with the conclusions stated in Theroem II by replacing $g+h>0$ by $h=g$ and $g+h<0$ by $h=-g$.

The question very naturally arises as to whether there exists a linear fractional transformation $u=(ex+f)/(gx+h)$ that will transform almost any distribution with the properties of $y=\psi(x)$ into a new distribution $v=\phi(u)$ with a mode at a previously assigned point $u=c$ within the range of admissible values of u. To insure a mode for $v=\phi(u)$ at $u=c$, it is, of course, sufficient that there exist values of e, f, g, and h that make the continuous function

$$(11)\quad \frac{dv}{du}=\frac{(he-fg)^2}{(gu-e)^4}\,\psi'\left(\frac{f-hu}{gu-e}\right)-\frac{2g(he-fg)}{(gu-e)^3}\,\psi\left(\frac{f-hu}{gu-e}\right)$$

change sign from positive to negative at $u=c$.

Since the only restrictions on e, f, g, and h are that

they shall be real, and that g and $he - fg$ shall be positive, it seems that the requirement that dv/du shall change from positive to negative at an assigned value of u could probably be satisfied for some important classes of relatively simple functions. As a simple example, take the quadratic function $\psi(x) = Ax^2 + Bx + C$, which, when subjected to the conditions on $\psi(x)$, becomes $\psi(x) = 3(1-x^2)/4$.

The mode is in this case at $x = 0$. The problem we propose is to find the linear fractional transformation $u = (ex+f)/(gx+h)$ that will transform $\psi(x)$ into $v = \phi(u)$ with a mode at an assigned $u = c$. In this case (11) becomes

$$\frac{dv}{du} = -\frac{3}{2}\frac{he-fg}{(gu-e)^5}\Big\{(he-fg)(f-hu) - \tag{12}$$
$$g\left[(g^2-h^2)u^2 + 2u(fh-eg) + e^2 - f^2\right]\Big\}.$$

To facilitate the examination of (12), make $h = g$. Then (12) reduces to

$$\frac{dv}{du} = -\frac{3}{2}\,\frac{g^2(e-f)^2}{(gu-e)^5}\,(e+2f-3gu). \tag{13}$$

Since $g + h > 0$, we have $gu - e < 0$, and consequently the coefficient of $(e+2f-3gu)$ is positive. To provide for the change of sign of (13) at $u = c$, select e, f, and g so that $e + 2f = 3cg$. To make (13) positive at $u = c - \delta$ and negative at $u = c + \delta$, where δ is arbitrarily small and positive, we may assign to g any positive value and to e any value greater than cg, for then f is less than e, which is the condition $he - fg > 0$ when $h = g$. While there are thus an infinite number of ways in which we may select a linear

fractional transformation so that, when applied to special functions, it will give a new distribution with a mode at an assigned point, no general proposition is proved that assures an assigned modal value of $\psi(x)$.

H. L. Rietz

ON SMALL SAMPLES FROM CERTAIN NON-NORMAL UNIVERSES*

By

PAUL R. RIDER
Washington University

INTRODUCTION

The distribution of the ratio

$$z = \frac{\text{mean of sample} - \text{mean of universe}}{\text{standard deviation of sample}}$$

which is of great importance in the theory of small samples, has been derived exactly by theoretical methods for samples of any size from a normal universe.[1] Experimental studies[2] have been

---

*The writer desires to express his grateful appreciation to the National Research Council, which made possible this study by a grant-in-aid for the assistance of a computer.

[1] See, for example, R. A. Fisher, Applications of "Student's" Distribution, Metron, vol. 5, No. 3 (Dec. 1, 1925), pp. 90-104. 5

[2] e. g. W. A. Shewhart and F. W. Winters, Small Samples—New Experimental Results, Journal of the American Statistical Association, Vol. 23 (1928), pp. 144-53;

J. Neyman and E. S. Pearson, On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. Part I, Biometrika, Vol. 20A (1928), pp. 175-240;

"Sophister," Discussion of Small Samples Drawn from an Infinite Skew Population, Biometrika, Vol. 20A (1928), pp. 389-423;

E. S. Pearson assisted by N. K. Adyanthāya and others, The Distribution of Frequency Constants in Small Samples from Non-normal Symmetrical and Skew Populations. 2nd paper, Biometrika, Vol. 21 (1929), pp. 259-86.

made of the z -distribution for samples of specific sizes from other types of universe. A theoretical method applicable to samples from a discrete universe was used in a previous paper,[1] in which a rectangular universe was studied in some detail. The rectangular universe was chosen as being the simplest from the standpoint of the method employed, and as a good example of a limited symmetric distribution. It is the purpose of the present paper to apply the method to a triangular population, which is a specimen of a limited skew distribution, and also to a U-shaped universe. The rectangular, triangular and U-shaped universes are shown in Table I in the columns headed R , T , and U , respectively. Their graphs are exhibited in Figure 1.

In addition to the z -distribution, the distributions of means from the triangular and from the U-shaped universe are given.

In the concluding section is discussed the probability corresponding to an interval of three sample standard deviations on each side of the sample mean.

All of the results of the paper are for samples of four.

THE DISTRIBUTION OF z

The distributions of z are shown in Table II,[2], in which the distribution for samples from a normal universe, N , is also given.

The cumulated probability of z for the triangular and for the U-shaped universe are shown in Table III, which may be compared with a similar table for a rectangular and for a normal universe given in Biometrika, Vol. 21 (1929), p. 131.

---

[1] P. R. Rider, On the Distribution of the Ratio of Mean to Standard Deviation in Small Samples from Non-normal Universes, Biometrika, Vol 21 (1929), pp. 124-143.

[2] For an explanation of the method of deriving these distributions see Rider, loc. cit.

These cumulated probabilities are plotted on probability paper in Figures 2 and 3 and may be compared with similar probabilities for a rectangular universe by reference to Biometrika, Vol. 21 (1929), p. 129, Figure 2.

The principal results to be noted are as follows:

1. The general characteristics of the z -distribution for the U-shaped universe are the same as those for a rectangular universe, viz. a greater number of z 's outside of a certain value of $|z|$, and also a greater clustering of z 's about the origin, than is the case for a normal universe.[1] This is to be expected, since the values of β_2 for U and R are 1.132 and 1.776 respectively, as compared with the value 3 for N .

2. The negative skewness in the triangular universe produces skewness of the opposite type in the distribution of z , as found experimentally by Neyman and E. S. Pearson[2] and by "Sophister."[3] This means (in the case of negative skewness in the universe) that the probability corresponding to an interval from $-\infty$ to z is smaller than when the sampling is from a normal universe.

3. The cumulated probability of $|z|$, or the probability corresponding to an interval from $-z$ to z , is somewhat the same for the triangular universe as for a normal universe;[4] a comparison is made in Table IV.

Results 2 and 3 are apparently due to the fact that in a

---

[1] See Rider, loc. cit., p. 130.

[2] Biometrika, Vol. 20A (1928), p. 198.

[3] Biometrika, Vol. 20A (1928), p. 408.

[4] cf. E. S. Pearson assisted by N. K. Adyanthāya and others. The Distribution of Frequency Constants in Small Samples from Non-normal Symmetrical and Skew Populations. 2nd paper, Biometrika, Vol. 21 (1929), pp. 259-86.

skew universe the regression of variance on mean[1] is often essentially linear (if parabolic, the vertex of the parabola is well to one side of the scatter diagram). Let us consider the case in which the slope of the regression line is positive. Designating by x the difference between the mean of a sample and the mean of the universe, and by s the standard deviation of the sample, we see that large values of $|x|$ tend to be associated with large values of s^2 (and therefore with large values of s). Thus the values of z tend to be smaller. On the other hand, for large values of $|-x|$, s is smaller and $|-z|$ consequently larger. This means that the frequencies corresponding to the algebraically lower values of z are greater than in the case of a normal universe, or that the use of "Student's" tables would give results too small for the probability that the mean of a sample does not exceed *algebraically* the mean of the universe by more than z times the standard deviation of the sample. The opposite is true in the case studied here, since the universe is negatively skew and the regression line of s^2 on x would have a negative slope.

Since there is a shifting of the whole cumulated z-distribution to the right or left, the effect noted in 3 is readily explained. As a result of this effect we should apparently not be far wrong, when sampling from a skew universe, if we used "Student's" tables to obtain the probability that the mean of a sample does not exceed *numerically* the mean of the universe by more than z times the standard deviation of the sample.[2]

---

[1] For the regression formula see J. Neyman, On the Correlation of the Mean and the Variance in Samples from an "Infinite" Population, Biometrika, Vol. 18 (1926), pp. 401-13.

[2] See E. S. Pearson assisted by N. K. Adyanthāya and others, The Distribution of Frequency Constants in Small Samples from Non-normal Symmetrical and Skew Populations. 2nd paper, Biometrika, Vol. 21 (1929), pp. 259-86.

THE DISTRIBUTION OF MEANS OF SAMPLES

The distributions of means of samples are shown in Tables V and VI. In these tables x indicates the difference between the mean of the sample and the mean of the universe.

For the difficulties involved in obtaining satisfactory results for the distribution of means of small samples from a U-shaped universe see K. J. Holtzinger and A. E. R. Church, "On the Means of Samples from a U-shaped Population," Biometrika, Vol. 20A (1928), pp. 361-88.

The probability corresponding to an interval of three sample standard deviations on each side of the sample mean.

If M is the mean and σ the standard deviation of a normally distributed variate X, then, as is well known, the probability that an item selected at random will lie within the range $M \pm 3\sigma$ is 0.997. If X and s are the mean and the standard deviation respectively of a sample, the expected or average probability corresponding to the interval $X \pm 3s$ will be different from the probability corresponding to the interval $M \pm 3\sigma$. Shewhart[1] obtained experimentally for the average probability for samples of four associated with the interval $X \pm 3s$ the values 0.90 for a normal universe, 0.91 for a rectangular universe, and 0.91 for a triangular universe.

By analyzing all possible samples of four from the rectangular and triangular universes of Table I it was possible to obtain the probability corresponding to an interval of $3s$ on either side of the sample mean. For example let us consider the sample (1, 1, 2, 2), for which $X = 1.5$, $s = 0.5$. The interval $X \pm 3s$ extends from 0 to 3. This interval includes 0.4 of the rectangular universe R; 0.4 then is the probability that an

[1] W. A. Shewhart, Note on the Probability Associated with the Error of a Single Observation, Journal of Forestry, Vol. 26 (1928) pp. 601-607.

observed value will fall within the interval. Now the particular sample (1, 1, 2, 2) would occur 6 times out of 10,000. If we take all of the samples for which the interval $X \pm 3s$ includes 0.4 of the rectangular universe we find that such samples occur 106 times out of 10,000. Such an analysis leads to Table VII, from which it is ascertained that the average probability corresponding to an interval of $X \pm 3s$ is 0.920. A similar analysis of the triangular universe T gives us Table VIII and yields 0.907 as the average probability associated with $X \pm 3s$. A better understanding of the situation may be obtained from Figure 4.

Paul R. Rider

TABLE I

Rectangular, Triangular and U-Shaped Universes

X	FREQUENCY		
	R	*T*	*U*
0	1		10
1	1	1	5
2	1	2	1
3	1	3	1
4	1	4	1
5	1	5	1
6	1	6	1
7	1	7	1
8	1	8	5
9	1	9	10
10		10	
Total	10	55	36
Mean	4.5	7	4.5
β_1*	0	0.326	0
β_2*	$1.7\dot{7}\dot{5}$	$2.3\dot{6}$	1.132+

*The values of the β's are uncorrected for grouping. The dots over the digits indicate repeating decimals. The values for a continuous rectangular distribution are $\beta_1 = 0$, $\beta_2 = 1.8$, and for a continuous triangular distribution are $\beta_1 = 0.32$, $\beta_2 = 2.4$.

TABLE II

Probability of z for Samples of 4

z	N	R	T	U
Below −4.25	.0026	.0077	.0015 +	.0384
−4.25 to −3.75	.0011	.0022	.0012	.0004
−3.75 to −3.25	.0018	.0026	.0007	.0009
−3.25 to −2.75	.0032	.0032	.0032	.0077
−2.75 to −2.25	.0062	.0074	.0028	.0016
−2.25 to −1.75	.0131	.0188	.0061	.0106
−1.75 to −1.25	.0314	.0267	.0251	.0147
−1.25 to −0.75	.0829	.0692	.0615	.0256
−0.75 to −0.25	.2047	.2000	.2098	.2299
−0.25 to 0.25	.3058	.3244	.3249	.3405 +
0.25 to 0.75	.2047	.2000	.1741	.2299
0.75 to 1.25	.0829	.0692	.0764	.0256
1.25 to 1.75	.0314	.0267	.0566	.0147
1.75 to 2.25	.0131	.0188	.0118	.0106
2.25 to 2.75	.0062	.0074	.0094	.0016
2.75 to 3.25	.0032	.0032	.0174	.0077
3.25 to 3.75	.0018	.0026	.0000	.0009
3.75 to 4.25	.0011	.0022	.0025 +	.0004
Above 4.25	.0026	.0077	.0150 −	.0383

TABLE III

The cumulated probability of z, or probability that the mean of a random sample of 4 will not exceed (in algebraic sense) the mean of the universe by more than z times the standard deviation of the sample.

z	Cumulated Probability Triangular Universe		Cumulated Probability U-Shaped Universe	
	for $-z$	for z	for $-z$	for z
0.0	.51955−	.51955−	.54355+	.54355+
.1	.41649	.54037	.39365−	.60635+
.2	.34497	.61053	.34651	.65349
.3	.28885+	.65136	.30739	.69261
.4	.22719	.70010	.27831	.72193
.5	.18568	.74269	.22081	.77991
.6	.14350−	.76942	.14785+	.85215−
.7	.11580	.79993	.11382	.88618
.8	.09485−	.81086	.09844	.90192
.9	.07784	.83462	.09065+	.90935+
1.0	.06130	.86748	.08285−	.91715+
1.1	.05053	.87456	.07994	.92006
1.2	.04256	.88731	.07471	.92529
1.3	.03716	.88731	.07363	.92637
1.4	.03152	.90787	.07179	.92821
1.5	.02783	.91316	.06614	.93387
1.6	.02334	.91911	.05979	.94021
1.7	.01845−	.93480	.05975−	.94025−
1.8	.01552	.94390	.05941	.94059
1.9	.01410	.94390	.05798	.94202
2.0	.01366	.94810	.05441	.94774
2.1	.01265−	.94810	.04959	.95041
2.2	.01039	.95565−	.04892	.95108
2.3	.00907	.95565−	.04892	.95108
2.4	.00871	.95565−	.04891	.95109
2.5	.00816	.95565−	.04891	.95118
2.6	.00725+	.95565−	.04803	.95197
2.7	.00725+	.95565−	.04732	.95268
2.8	.00661	.96509	.04728	.95272
2.9	.00483	.97910	.04133	.95867
3.0	.00462	.98250−	.03954	.96046
3.5	.00272	.98250−	.03904	.96132
4.0	.00242	.98250−	.03833	.96168

TABLE IV

Cumulated Probability of $|z|$ for Samples of 4.

$\lvert z \rvert$ greater than	Probability		$\lvert z \rvert$ greater than	Probability	
	Triangular Universe	Normal Universe		Triangular Universe	Normal Universe
0.0	.9219	1.0000	1.6	.1042	.0695−
.1	.8761	.8735+	1.7	.0836	.0603
.2	.7303	.7519	1.8	.0716	.0526
.3	.6375−	.6392	1.9	.0702	.0460
.4	.5271	.5382	2.0	.0652	.0405+
.5	.4423	.4502	2.1	.0646	.0358
.6	.3723	.3751	2.2	.0547	.0318
.7	.3135−	.3121	2.3	.0534	.0283
.8	.2834	.2599	2.4	.0531	.0253
.9	.2432	.2169	2.5	.0525+	.0227
1.0	.1891	.1817	2.6	.0516	.0204
1.1	.1755−	.1528	2.7	.0516	.0185−
1.2	.1552	.1292	2.8	.0415+	.0167
1.3	.1497	.1098	2.9	.0257	.0152
1.4	.1236	.0938	3.0	.0212	.0138
1.5	.1146	.0805+			

TABLE V

Distribution of Means of Samples of 4 from Triangular Universe

x	Probability	x	Probability	x	Probability
–5.25	.00001	–2.25	.01627	0.75	.07202
–5.00	.00004	–2.00	.02200	1.00	.06437
–4.75	.00009	–1.75	.02882	1.25	.05496
–4.50	.00019	–1.50	.03559	1.50	.04462
–4.25	.00038	–1.25	.04501	1.75	.03415+
–4.00	.00070	–1.00	.05362	2.00	.02430
–3.75	.00125	–0.75	.06187	2.25	.01569
–3.50	.00212	–0.50	.06916	2.50	.00881
–3.25	.00344	–0.25	.07484	2.75	.00393
–3.00	.00537	0.00	.07834	3.00	.00109
–2.75	.00805–	0.25	.07918		
–2.50	.01165–	0.50	.07707		

x = (mean of sample) – (mean of universe)

TABLE VI

Distribution of Means of Samples of 4 from U-Shaped Universe

x	Frequency	Probability	x	Frequency	Probability
–4.50	10000	.0060	0.25	106660	.0635+
–4.25	20000	.0119	0.50	62755	.0374
–4.00	19000	.0113	0.75	51244	.0305+
–3.75	15000	0089	1.00	49270	.0293
–3.50	14225	.0085–	1.25	48376	.0288
–3.25	15300	.0091	1.50	49505	.0295–
–3.00	16690	.0099	1.75	63960	.0381
–2.75	18140	.0108	2.00	89660	.0534
–2.50	35651	.0212	2.25	81224	.0484
–2.25	81224	.0484	2.50	35651	.0212
–2.00	89660	.0534	2.75	18140	.0108
–1.75	63960	.0381	3.00	16690	.0099
–1.50	49505	.0295–	3.25	15300	.0091
–1.25	48376	.0288	3.50	14225	.0085–
–1.00	49270	.0293	3.75	15000	.0089
–0.75	51244	.0305+	4.00	19000	.0113
–0.50	62755	.0374	4.25	20000	.0119
–0.25	106660	.0635+	4.50	10000	.0060
0.00	146296	.0871			
			Total	1679616	1.0001

x = (mean of sample) – (mean of universe)

TABLE VII

Probability Corresponding to the Interval $\bar{X} \pm 3s$
Rectangular Universe

Proportion of universe included in $\bar{X} \pm 3s$*	Number of samples for which this proportion occurs**
0.1	10
0.2	8
0.3	84
0.4	106
0.5	284
0.6	324
0.7	564
0.8	652
0.9	888
1.0	7080
Total	10000

[1] i. e. the probability corresponding to $\bar{X} \pm 3s$.

[2] The probability of the occurrence of this proportion is, of course, obtained by dividing by 10000.

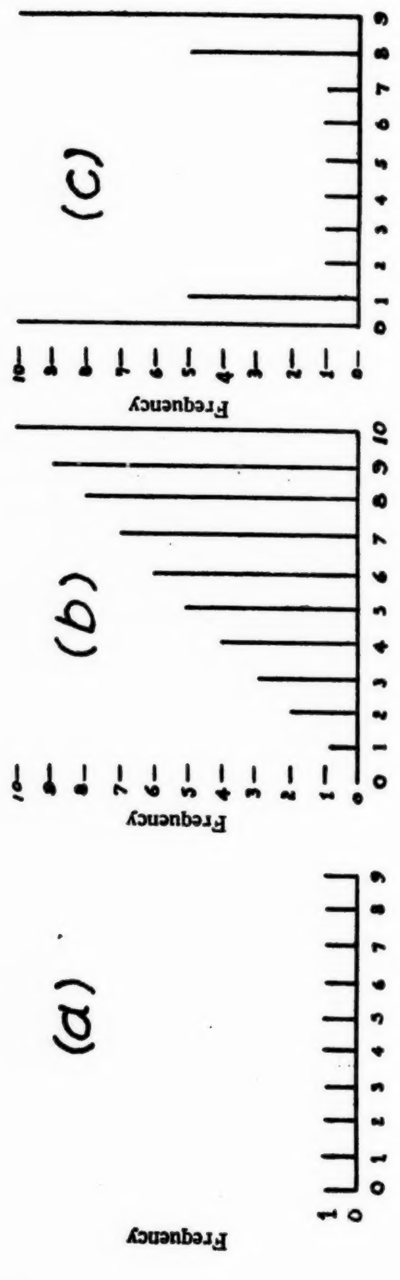
TABLE VIII

Probability Corresponding to the Interval $X \pm 3s$

Triangular Universe

Proportion of universe included in $\bar{X} \pm 3s$	Number of samples for which this proportion occurs	Probability of occurrence of this proportion	Cumulated probability
1/55 = .018	1	—	—
2/55 = .036	16	—	—
3/55 = .055–	89	—	—
4/55 = .073	256	—	—
5/55 = .091	625	.0001	.0001
6/55 = .109	1448	.0002	.0003
7/55 = .127	2401	.0003	.0006
8/55 = .145–	4096	.0004	.0010
9/55 = .164	6993	.0008	.0018
10/55 = .182	11388	.0012	.0030
12/55 = .218	1280	.0001	.0031
13/55 = .236	7776	.0008	.0039
14/55 = .255–	2928	.0003	.0042
15/55 = .273	8762	.0010	.0052
18/55 = .327	12768	.0014	.0066
19/55 = .345+	36000	.0039	.0105
20/55 = .364	8640	.0009	.0114
21/55 = .382	26508	.0029	.0143
22/55 = .400	5400	.0006	.0149
24/55 = .436	32768	.0036	.0185
25/55 = .455–	21600	.0024	.0209
26/55 = .473	10584	.0012	.0221
27/55 = .491	112764	.0123	.0344
28/55 = .509	19698	.0022	.0366
30/55 = .545+	71526	.0078	.0444
33/55 = .600	27116	.0030	.0474
34/55 = .618	296384	.0324	.0798
35/55 = .636	115128	.0126	.0924
36/55 = .655–	37892	.0041	.0965
39/55 = .709	54092	.0059	.1024
40/55 = .727	555924	.0608	.1632
42/55 = .764	57888	.0063	.1695
44/55 = .800	26416	.0029	.1724
45/55 = .818	556520	.0608	.2332
49/55 = .891	774320	.0846	.3178
52/55 = .945+	904676	.0989	.4167
54/55 = .982	879564	.0961	.5128
55/55 = 1.000	4458390	.4872	1.0000
Total	9150625	1.0000	

i. e. the probability corresponding to $\bar{X} \pm 3s$



Rectangular Universe

Triangular Universe

U-Shaped Universe

FIGURE 1

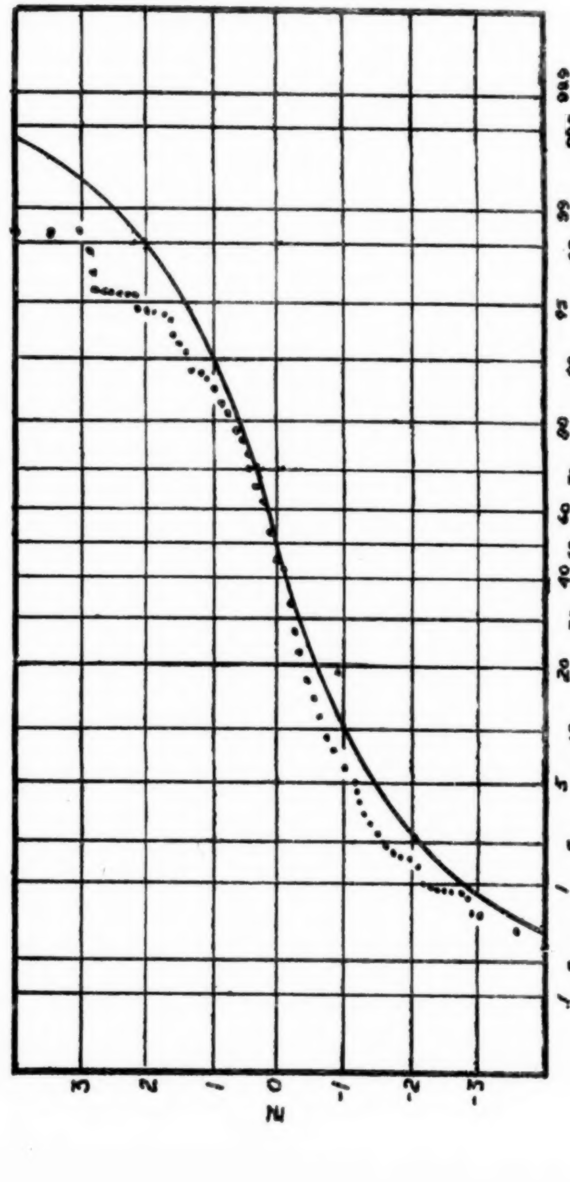


FIGURE 2

Cumulated Probability of z —Triangular Universe
The curve is for samples of 4 from a normal universe.
The dots are for samples of 4 from the universe T .

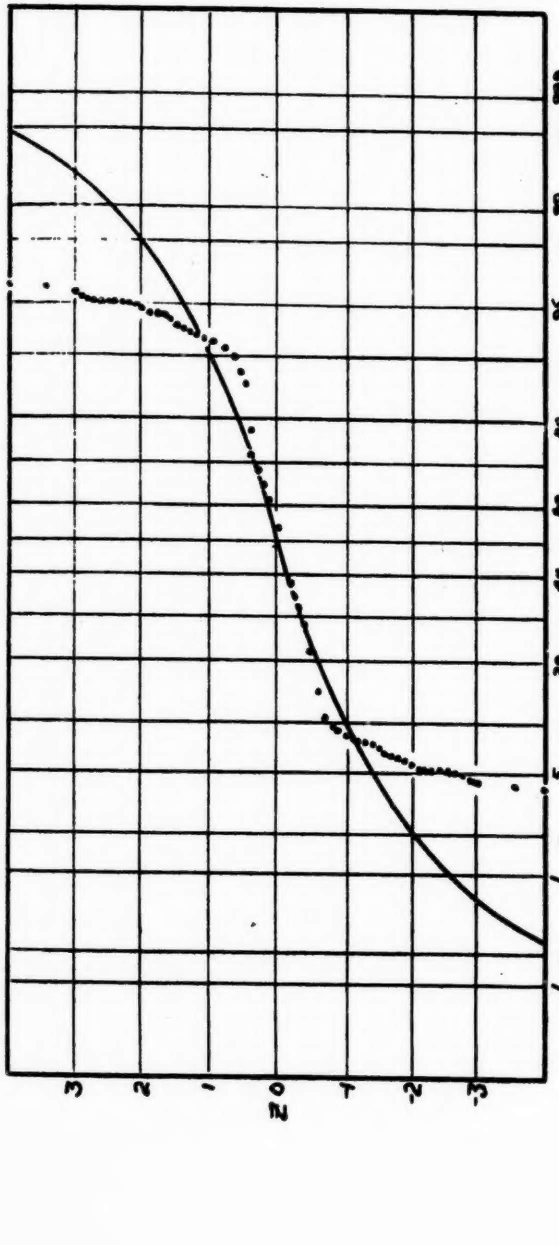


FIGURE 3

Cumulated Probability of z —U-Shaped Universe
The curve is for samples of 4 from a normal universe.
The dots are for samples of 4 from the universe U.

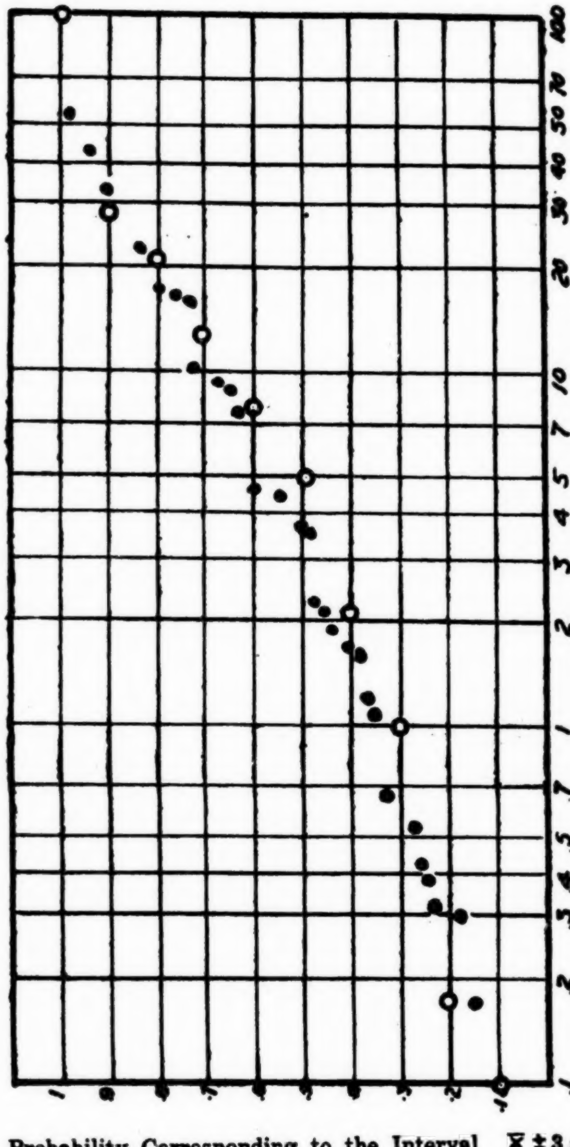


FIGURE 4

Probability Corresponding to the Interval $\bar{X} \pm 3s$
The circles are for samples of 4 from a rectangular universe, the dots for samples of 4 from a triangular universe.

AN EMPIRICAL DETERMINATION OF THE DISTRIBUTION OF MEANS, STANDARD DEVIATIONS AND CORRELATION COEFFICIENTS DRAWN FROM RECTANGULAR POPULATIONS*

By

HILDA FROST DUNLAP

Territorial Normal and Training School, Honolulu, Hawaii

Formulae for the standard errors of means, standard deviations and correlation coefficients have been derived on the assumption of a normal distribution in the sampled population. They are said to serve approximately even when the population varies considerably from the normal. This paper presents empirical evidence of their applicability in the case of means and standard deviations of samples of ten from a rectangular discontinuous population, and of correlation coefficients of samples of fifty-two from a rank distribution.

The data for the study of the distribution of means and standard deviations were secured by throwing ten dice 1600 times.

The dice were cubes four-tenths of an inch along an edge and numbered on opposite faces 1-6, 2-5, 3-4. They were constructed of bone and formed a matched set.

*The writer is indebted to Jack W. Dunlap for reading the entire manuscript and for checking the mechanical computations.

These were thrown from a cup whose inside diameter was 1.75 inches and whose depth was 2.5 inches. The dice were shaken in a box and then cast upon an especially prepared flat topped table covered with eight thicknesses of an army blanket.

As a guard against any possible bias in the table, the dice were thrown alternately with the right and left hands. After each throw the number of aces, deuces, treys, fours, fives, and sixes were recorded, and the mean and standard deviation calculated. In this study each throw was taken as a sample of ten drawn from a population of 16,000.

The next step was to determine whether there was any systematic bias in the dice used. The *a priori* expectation for any particular face of the die is one-sixth, here one sixth of 16,000, or 2,666⅔. This is of the nature of a point binomial of the form $(p+q)^n$ with a standard deviation equal to $\sqrt{Npq}$

TABLE I

Distribution of Observed and Theoretical Populations with a Test of the Difference of Their Standard Deviations

Die Face	Observed Frequency	Expected Frequency	Difference
1	2726	2666⅔	59⅓
2	2653	2666⅔	14⅔
3	2671	2666⅔	4⅓
4	2763	2666⅔	96⅓
5	2650	2666⅔	17⅔
6	2537	2666⅔	130⅔

$\sigma = (1600 \cdot 1/6 \cdot 5/6)^{\frac{1}{2}} = 47.1$ $s = (\sum d^2/N)^{\frac{1}{2}} = 70.8$

$s - \sigma = 23.7 \pm 13.76$

Table I gives the observed and expected values of each face. The standard deviation of the differences was determined and compared with the standard deviation of the expected distribution and the probable error of this difference was found.

Small s is used here to denote a standard deviation of a sample, while σ represents the standard deviation of the theoretical or true population. The formula for the standard deviation of a difference is

$$\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2 r_{12} \sigma_1 \sigma_2}$$

and in particular

$$\sigma_{s-d} = \sqrt{\sigma_s^2}$$

The second term drops out here because it is the standard deviation of the true standard error and this is equal to zero. The third term drops out for the same reason. Table I shows that the difference between the obtained and expected standard deviations is 23.7 ± 13.76. As this is less than twice its probable error, it can be concluded that the difference is not significant and that there is no significant bias in the dice.

MEANS

Figure 1 shows the distribution of the 1600 observed means. a normal curve for $N = 1600$ is superimposed on the histogram. For this distribution

$$r_1 \; (= \sqrt{\beta_1} \,) = .0160 \pm .0413, \text{ indicating symmetry}$$

$$r_2 \; (\sqrt{\beta_2} - 3 \,) = -.1050 \pm .0826, \text{ indicating mesokurtosis}$$

whence we may conclude that the normal curve represents this

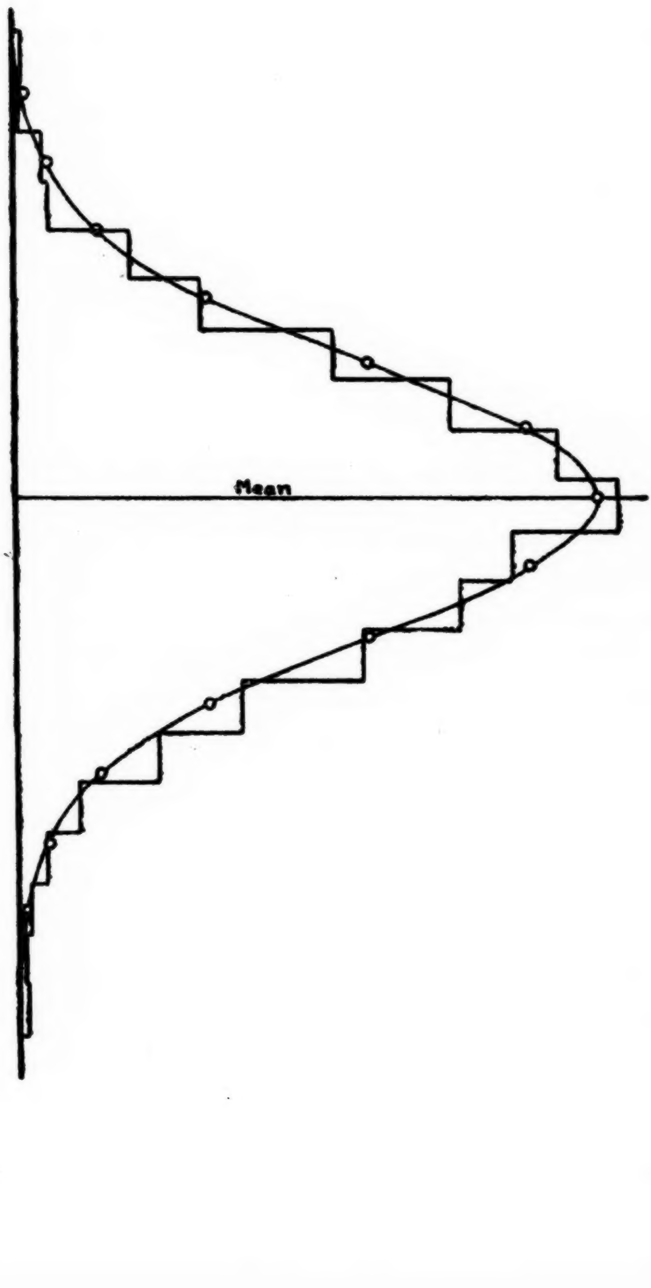


FIGURE 1
Distribution of 1600 means of samples of ten, with fitted normal curve.

distribution adequately.

The curves of this and succeeding figures were drawn through points calculated at intervals of ½ σ, except that in the case of Figures 2 and 3, points beyond ± 2 σ were calculated at intervals of 1 σ.

The values of the observed means varied from 1.6 to 5.4, a range of 6.9129 standard deviations.

The basic information to be drawn from this study of the distribution of 1600 means of samples of ten is given in Table II. The table is interpreted as follows:

The mean of the sampled population (16,000) is 3.47306, while the theoretical mean of the infinite population is 3.500000. The standard deviation of the sampled population (16,000) is 1.6788, and of the theoretical population 1.7078. The standard error of the mean of the sampled population is .0133. In comparing the mean of the sampled population with the mean of the theoretical infinite population, the former is treated as an experimental value whose standard error can be estimated, while the latter, being a true value, has no error.

The standard deviation of the difference between the means M (theoretical population) and $\bar{x}$ (sampled population) is

$$\sigma_{(M-\bar{x})} = \sqrt{\sigma_M^2 + \sigma_{\bar{x}}^2 - 2\,r_{M\bar{x}}\,\sigma_M\,\sigma_{\bar{x}}}$$

$$\sqrt{\sigma_{\bar{x}}^2} = .0133$$

The first and third terms drop out because σ_M equals zero. The difference between the mean of the theoretical population and the sampled population is .02694 ± .00897, from which it can be concluded that the mean tends to vary from the true mean.

$\bar{x}$ will hereafter refer to the mean of a sample of ten. The best estimate of the mean of a sample of ten that can be made for any sample chosen at random from the sampled population

TABLE II

Distribution of 1600 Means of Samples of 10

Description	Observed Value ($\bar{x}$)	Theoretical Value (M)
Mean of Sampled Pop.	3.47306	3.5000
σ of Sampled Pop.	1.6788	1.7078
σ_{mean} of Sampled Pop.	.0133	.0000
$\sigma_{(M-\bar{x})}$ of Sampled Pop.	.0133	.0000
$M-\bar{x}$ of Sampled Pop.	.0269 ± .00897	.0000
Mean of Means of Samples	3.47306	3.47306 or 3.5000
S. D. of Means of Samples	.5497	.5372 or .5401
S. E. of S. D. of Means of Samples	.0097	.0000 or .0000
$s_{\bar{x}} - \sigma_M$	.0125 ± .0065 or .0096 ± .0065	.0000 or .0000
γ_1 of Distri. of Means of Samples	.0160 ± .0413	.00 (normal theory)
γ_2 of Distri. of Means of Samples	−.1050 ± .0826	.00 (normal theory)

is 3.47306, and from the infinite population, 3.5000.

The standard deviation of the means of 1600 samples is .5467, while the estimated value for a sample picked at random from the sampled population is .5372 and from the theoretical infinite population .5401. These last two values are calculated by the formula

The best estimate of the standard deviation of a sample of ten picked at random from the sampled population is the σ of the sampled population, 1.6788, or of the theoretical infinite population, 1.7078, whence the values in the tables are obtained.

The standard error of the standard deviation of the means of samples is .0097. The standard error of the standard error σ_M of the mean of a sample of ten from the sampled and theoretical infinite populations is zero, as these are true values.

The difference between the standard deviation of the means and the standard error of such means of samples of ten from the sampled population or the theoretical infinite population is .0125 ±.0065. Thus there is no significant difference between the value of σ_M when calculated by the formula $\sigma_M = \frac{\sigma}{\sqrt{N}}$ and an actual distribution when samples as small as ten are used.

γ_1 indicates, as pointed out above, that the distribution is not skewed, while γ_2 shows the distribution to be slightly peaked but not significantly so.

STANDARD DEVIATIONS

Figure 2 shows a histogram and a fitted Gram-Charlier Type A curve, of the distribution of 1600 standard deviations of samples of ten calculated by the formula

$$s = \sqrt{\frac{\sum x^2}{N}}$$

X being measured from the mean, $\bar{x}$.

Figure 3 shows a similar histogram and curve fitted to the

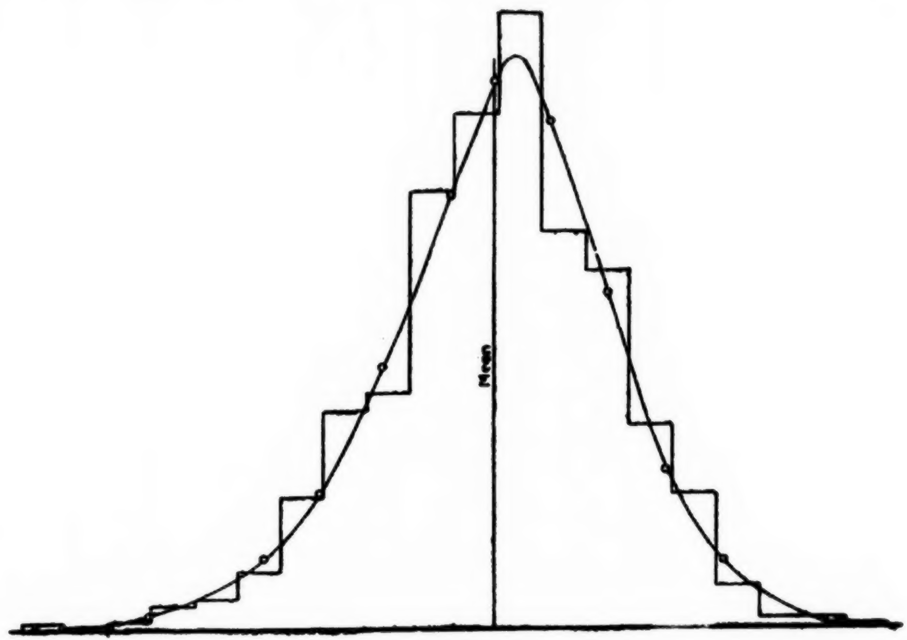


FIGURE 2

Distribution and fitted Gram-Charlier curve of 1600 standard deviations of samples of ten, calculated by the formula $s = \left(\frac{1}{N} \Sigma x^2 \right)^{\frac{1}{2}}$

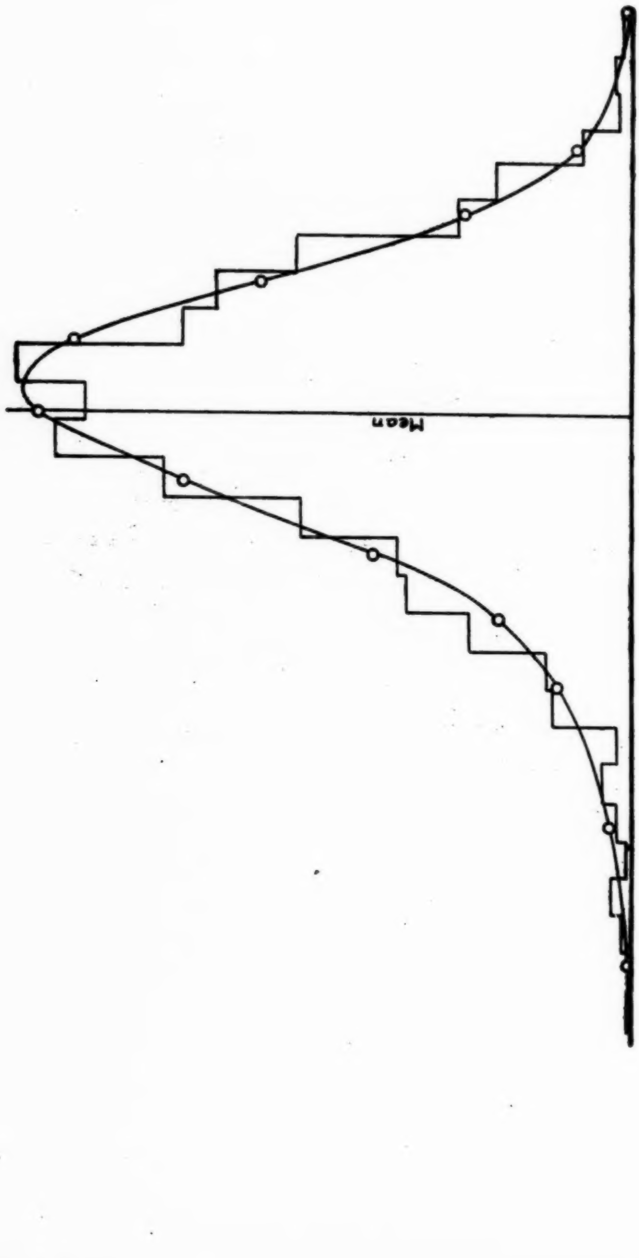


FIGURE 3

Distribution and fitted Gram-Charlier curve of 1600 standard deviations of samples of ten, calculated by the formula $s=\left(\frac{1}{N-1}\sum x^2\right)^{\frac{1}{2}}$

TABLE III

Distribution of 1600 Standard Deviations of Samples of Ten

Description	Observed Value		Theoretical Value	
	$s^2 = \frac{\Sigma x^2}{N}$	$s^2 = \frac{\Sigma x^2}{N-1}$	Sampled Population	Infinite Population
$\bar{x}$ of s's of sam.	1.5869	2.0403	1.6988	1.7078
S. D. of s's of sam.	.2665	.2538	.3799	.3818
S. D. of $\bar{x}$ of s's of samples	.0067	.0063	.0000	.0000
S. D. of s of s's of samples	.0047	.0045	.0000	.0000
$\sigma - \bar{x}_s$	.1119 ±0045 or .1209 ±.0045	.3415 ±.0042 or .3325 ±.0042	.0000	.0000
$\sigma_\sigma - s_s$	.1134 ±.0032 or .1153 ±0032	.1261 ±.0030 or .1280 ±.0030	.0000	.0000
γ_1 (skewness)	−.3568 ±0413	−.5026 ±.0413	.0000 (normal theory)	
γ_2 (kurtosis)	.5140 ±0826	.6851 ±.0826	.0000 (normal theory)	
N	1600	1600		

same data when the standard deviations are calculated by the formula

$$s = \sqrt{\frac{\Sigma x^2}{N-1}}$$

A study of this latter formula is included here to test which is more appropriate when dealing with small samples from a rectangular population.

An interpretation of Table III is now in order. Column one is a description of the statistics involved. Column two is subdivided into two parts: First, when s equals $\sqrt{\frac{\Sigma x^2}{N}}$, and second when s equals $\sqrt{\frac{\Sigma x^2}{N-1}}$. Column three gives the theoretical values. There are two of these—one for the sampled population and one for the infinite population. In the case of the sampled population the values calculated for the standard deviation and the σ_σ become true values when a single sample is compared with them in exactly the same manner as if compared with similar values from the infinite population. The reason for this is that for a given sample the 16,000 constitutes the actual population from which the sample is drawn.

In the first line the means of the standard deviations of the samples are found to equal respectively, 1.5869 and 2.0403. The theoretical means for the sampled and infinite populations are respectively 1.6988 and 1.7078.

In the next line are the standard deviations of standard deviations of samples. These are calculated values, obtained by substituting in the formula

$$\sigma_s = \frac{\sigma}{\sqrt{2N}}$$

As the best estimate of the standard deviations of any particular sample chosen at random is the standard deviation of the sampled population, or the infinite population, these values can be substituted in the above formula in obtaining the standard error of the standard deviation of such a sample of ten.

The standard error of the mean of standard deviations in

samples for both observed values is given in line three. Obviously in the case of the sampled and infinite populations these equal zero. It should be clearly understood by the reader that here N equals 1600, the number of standard deviations used in determining the mean standard deviation.

Line four gives the standard error of the standard deviation of standard deviations of samples of ten.

Line five gives the difference between each of the true standard deviations (sampled and infinite) and the two observed mean standard deviations. The standard deviations of the sampled population and of the infinite population are each greater than the mean standard deviation of the observed population when calculated by the formula $s=\sqrt{\frac{\Sigma x^2}{N}}$. In the first case the difference is .1119 ± .0045. This is approximately 25 times its probable error, so it must be considered a significant difference. The difference when compared with the theoretical infinite population is .1209 ± .0045. This is even more significant. When the theoretical values are compared with the mean standard deviation calculated by the formula $s=\sqrt{\frac{\Sigma x^2}{N-1}}$ the differences are found to be .3415 ± .0042, and .3325 ± .0042. The differences here are much greater than those found from the first formula.

Line six shows the difference between the standard errors of the standard deviations of the true populations and the calculated s_s of the samples. The difference between σ_σ and s_s (.3799 −.2665), is .1134 ± .0032. This difference is approximately 35 times its probable error. The difference between .3799 and .2538 is even greater. Still larger differences are found when s_s is calculated for the $s=\sqrt{\frac{\Sigma x^2}{N-1}}$ formula.

γ_1 in the case of both curves is negative and more than 8 times its probable error, definitely showing a negative skewness. γ_2 in the case of both curves is 6 times greater than its probable error, indicating definite leptokurtosis. The Gram-Charlier curves shown in Figures 2 and 3 were fitted to the first four

moments according to the equation

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[1 - \left(\frac{\mu_3}{6\sigma^3}\right)(3x - x^3) + \left(\frac{\mu_4}{24\sigma^4} - .125\right)(x^4 - 6x^2 + 3) \right]$$

where $$x = \frac{X - \bar{x}}{s}$$

If we compute values of s by the empirical formula $s = \sqrt{\frac{\Sigma x^2}{N - .25}}$, the mean value is 1.7039, which lies very close to the theoretical values 1.6988 and 1.7078, in fact almost exactly half-way between them.

CORRELATION COEFFICIENTS

The product-moment correlation coefficient varies between the limits plus one and minus one. Obviously, the distribution of correlation coefficients cannot be normal, although in the case where $r = 0$ their distribution should approximate a normal curve, as it can become symmetrical. Coefficients around any other point tend to be distributed asymmetrically.

It was assumed that if a deck of cards be thoroughly shuffled there should be no correlation between successive deals. Using a deck of cards gives a sample of 52. A new pack was thoroughly shuffled. The cards were then dealt one at a time, the first card dealt being recorded as number one, the second card dealt as number two, the third card as number three, etc. That is, if the seven of hearts was turned first, the value one was recorded against its place in the table. After each deal the cards were picked up in the same order and shuffled three times by the fan method and then cut twice. Sixty such deals were made and recorded. Then rank correlations were calculated be-

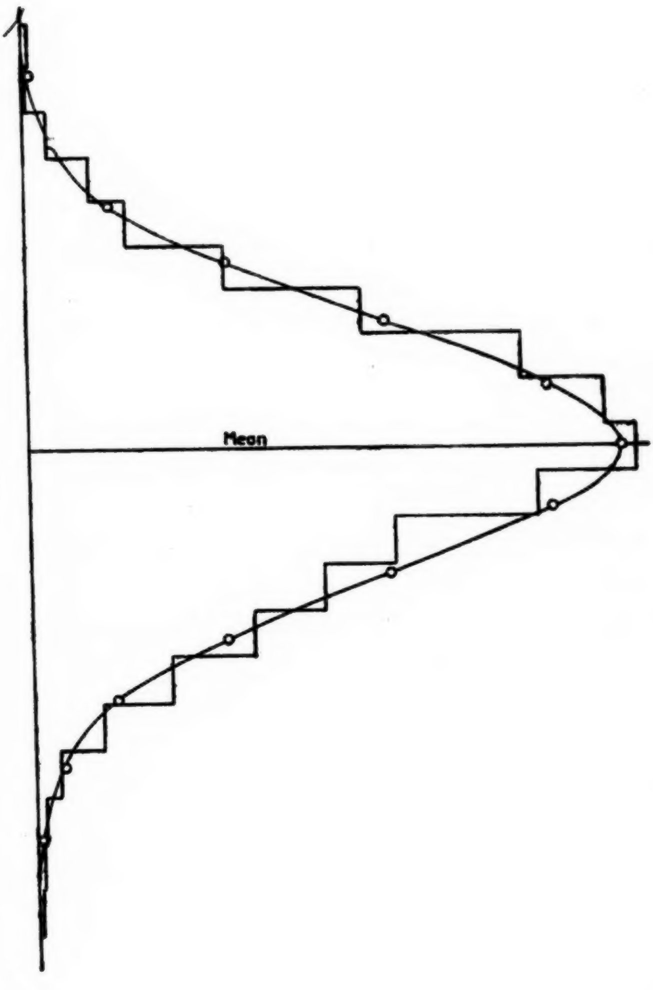


FIGURE 4

Distribution of 1770 correlation coefficients of samples of 52, with fitted normal curve.

tween each pair of deals, the total number of intercorrelations being $\frac{n(n-1)}{2}$, here 1770.

In this study, there could be no split ranks. Each card could receive one and only one rank on each deal. Thus, the rank correlation formula gave exactly the same values as would a Pearson product-moment coefficient.

Figure 4 shows a histogram with a fitted normal curve superimposed on it. γ_1 for this curve is .000015 ± .0392, indicating no skewness, and γ_2 is .2174 ± .0785, indicating a slight tendency to peakedness. Both of these facts are shown by the fit of the curve to the histogram.

The formula for the standard error of a correlation coefficient from a normal population is

$$\sigma_r = \frac{1-\rho^2}{\sqrt{N}}$$

ρ being the correlation in the population. Thus when $r = .0000$ and $N = 52$, $\sigma_r = .1387$.

The mean value of the 1770 coefficients is $r = -.0012$. The expected mean is zero. The difference between these two values is .0012 ± .0022. This shows that the mean correlation coefficient is not significantly different from the expected mean correlation.

The standard deviation of the observed distribution is .1359. This value differs from the expected value by .0028 ± .0091. The formula $\sigma_r = \frac{1-\rho^2}{\sqrt{N}}$ is therefore seen to give a sufficiently close approximation in this case.

CONCLUSIONS

1. The distribution of means of samples of ten drawn from a discontinuous rectangular population is normal. The formula $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ gives a reasonably close estimate of the standard error of such means.

2. The distribution of standard deviations of samples of

ten drawn from a discontinuous rectangular population is skewed and leptokurtic. The formula $\sigma_s = \frac{\sigma}{\sqrt{2N}}$ does not give a reasonably close estimate of the standard deviation of standard deviations of samples of ten, whether the latter are computed from the formula $s = \sqrt{\frac{\Sigma x^2}{N}}$ or $s = \sqrt{\frac{\Sigma x^2}{N-1}}$

3. Neither of the formulas, $s = \sqrt{\frac{\Sigma x^2}{N}}$ and $s = \sqrt{\frac{\Sigma x^2}{N-1}}$ for the standard deviation of a sample of ten gives a reasonably close estimate of the true standard deviation in a rectangular discontinuous population. The empirical formula $s = \sqrt{\frac{\Sigma x^2}{N-.25}}$ does appear to do so.

4. The distribution of correlation coefficients of samples of 52 from a rank population in which the expected correlation is zero, is symmetrical and very slightly leptokurtic. The formula $\sigma_r = \frac{1-\rho^2}{\sqrt{N}}$ represents adequately the standard deviation of such correlation coefficients.

Hilda F. Dunlap

EDITORIAL

The Interdependence of Sampling and Frequency Distribution Theory

The object of the theory of sampling is to describe the phenomena exhibited by all the samples that can possibly arise from a parent population of known characteristics. In some cases the desired description can be obtained directly by employing elementary operations of combination theory, in others it is either expedient or necessary to use the indirect attack of the statistical theory of sampling. These two methods are quite different in application, and it is advisable to illustrate the respective peculiarities of the two methods.

Example 1. An auction bridge hand may be regarded as a single sample withdrawn from a parent population of 52 cards. The number of different hands that can be selected equals the number of combinations of 52 things taken 13 at a time, namely, $\binom{52}{13} = 635\ 013\ 559\ 600$. Of these

$$(1) \cdots\cdots\cdots \quad f(z) = \binom{39}{13-z}\binom{13}{z}$$

will contain exactly z cards of any specified suit. Therefore if in this expression we successively place z equal to 0, 1, 2, . . . 13 we shall obtain the frequency of all possible samples ranked according to the number of cards of the specified suit contained in each sample. The results are presented in the following table.

TABLE I

z	$f(z)$	$P_z = f(z)/N$
0	8 122 425 444	.01279
1	50 840 366 668	.08006
2	130 732 371 432	.20587
3	181 823 183 256	.28633
4	151 519 319 380	.23861
5	79 181 063 676	.12469
6	26 393 687 892	.04156
7	5 598 661 068	.00882
8	740 999 259	.00117
9	58 809 465	.00009
10	2 613 754	.00000
11	57 798	.00000
12	507	.00000
13	1	.00000
Total	635 013 559 600	.99999

In this illustration, combination theory has yielded a perfect solution. The frquencies are exact, and the sum of the frequencies between any two limits may likewise be obtained exactly by a simple addition.

Example 2. The bidding strength of hands in auction bridge is often approximated by counting each Jack, Queen, King and Ace as 1, 2, 3 and 4 points, respectively. The total count of a single hand may range, therefore from 0 to 37 inclusive. Required the frequency distribution of all possible hands when they

are classified according to count.

Unlike the preceding problem, we cannot obtain a simple expression for the general term, f_z, of the required distribution. But after rather involved computations the following solution may be obtained:

TABLE II

Count z	Frequency $f(z)$	Count z	Frequency $f(z)$
0	2 310 789 600	19	6 579 838 440
1	5 006 710 800	20	4 086 538 404
2	8 611 542 576	21	2 399 507 844
3	15 636 342 960	22	1 333 800 036
4	24 419 055 136	23	710 603 628
5	32 933 031 040	24	354 993 864
6	41 619 399 184	25	167 819 892
7	50 979 441 968	26	74 095 248
8	56 466 608 128	27	31 157 940
9	59 413 313 872	28	11 790 760
10	59 723 754 816	29	4 236 588
11	56 799 933 520	30	1 396 068
12	50 971 682 080	31	388 196
13	43 906 944 752	32	109 156
14	36 153 374 224	33	22 360
15	28 090 962 724	34	4 484
16	21 024 781 756	35	624
17	14 997 080 848	36	60
18	10 192 504 020	37	4
		Total	635 013 559 600

Example 3. If the mean and the standard deviation of the weights of a group of 200,000 men be 140 lbs. and 20 lbs., respectively, and if in addition it be known that the higher standard moments of this distribution be

$$\alpha_{3:x} = .5 \qquad \alpha_{5:x} = 4.43$$

$$\alpha_{4:x} = 3.17 \qquad \alpha_{6:x} = 17.97,$$

what is the chance that the mean weight of 1000 men chosen at random from the 200,000 will exceed 141 pounds?

It is clear that it would be physically impossible to solve this problem by employing a direct attack by combination theory, even though the weights of each of the 200,000 men were available. Moreover, it is likewise evident that in statistical problems corresponding to the illustrations of examples 1 and 2, the number of individuals in both the parent population and each sample is considerably larger than 52 and 13 respectively, and consequently the calculation of either a single frequency or the sum of any large group of consecutive frequencies by the direct method is quite out of the question.

Let us now consider the three examples above from the point of view of the indirect attack. The parent populations for the first two examples may be interpreted as

Variates . . .	x	0	1
Frequencies . .	$f(x)$	39	13

and

Variates . .	x . .	0	1	2	3	4
Frequencies . .	$f(x)$. .	36	4	4	4	4

respectively.

For the first, the mean is at $x = 1/4$, and the moments about the mean of the parent population are obviously

$$\mu_{n:x} = \frac{13}{4^n}\left[3^n + 3(-1)^n\right]$$

For the second, the mean is at $x = 10/13$, and correspondingly the moments of this parent population are

$$\mu_{n:x} = \frac{1}{13^n}\left[(-10)^n + 3^n + 16^n + 29^n + 42^n\right]$$

If s and r denote the number of individuals in the parent population and each sample respectively, then the moments of the distribution of all samples that can arise from this parent population may be obtained from those of the parent population by means of the relations

$$(2)^1 \left\{ \begin{aligned} & M_z = r \cdot M_x \\ & \mu_{2:z} = \mu_{2:x} \cdot s(\rho_1 - \rho_2) \\ & \mu_{3:z} = \mu_{3:x} \cdot s(\rho_1 - 3\rho_2 + 2\rho_3 \\ & \mu_{4:z} = \mu_{4:x} \cdot s(\rho_1 - 7\rho_2 + 12\rho_3 - 6\rho_4) + 3\mu_{2:x}^2 \cdot s^2(\rho_2 - 2\rho_3 + \rho_4) \\ & \mu_{5:z} = \mu_{5:x} \cdot s(\rho_1 - 15\rho_2 + 50\rho_3 - 60\rho_4 + 24\rho_5) \\ & \qquad + 10\mu_{3:x}\,\mu_{2:x} \cdot s^2(\rho_2 - 4\rho_3 + 5\rho_4 - 2\rho_5) \\ & \mu_{6:z} = \mu_{6:x} \cdot s(\rho_1 - 31\rho_2 + 180\rho_3 - 390\rho_4 + 360\rho_5 - 120\rho_6) \\ & \qquad + 15\mu_{4:x} \cdot \mu_{2:x} \cdot s^2(\rho_2 - 8\rho_3 + 19\rho_4 - 18\rho_5 + 6\rho_6) \\ & \qquad + 10\mu_{3:x}^2 \cdot s^2(\rho_2 - 6\rho_3 + 13\rho_4 - 12\rho_5 + 4\rho_6) \\ & \qquad + 15\mu_{2:x}^3 \cdot s^3(\rho_3 - 3\rho_4 + 3\rho_5 - \rho_6) \end{aligned} \right.$$

where

$$\rho_i = \frac{r(r-1)(r-2)\cdots \text{ to } i \text{ factors}}{s(s-1)(s-2)\cdots \text{ to } i \text{ factors}}$$

Since the moments $\mu_{n:x}$ for each of these three examples are now known, and according to the conditions of the problems the values of (r, s) are (13, 52), (13, 52), and (1000, 200000) respectively, it follows that the moments of the desired distributions of samples are as follows:

Function	Example 1	Example 2	Example 3
M_x	13/4	10	$M_x = 140$ lbs.
$\mu_{2:x}$	507/272	290/17	$\sigma_x =$ .630874 lbs.
$\mu_{3:x}$	6591/13600	288/17	$\alpha_{3:x} =$ .0156927
$\mu_{4:x}$	53591421/5331200	17441114/29155	$\alpha_{4:x} =$ 3.0001357
$\mu_{5:x}$	9339447/1066240	2262240/833	$\alpha_{5:x} =$ .1569051
$\mu_{6:x}$	71781968037/801812480	2684384074/39151	$\alpha_{6:x} =$ 15.026638

It will be observed that the indirect procedure has yielded the moments of the required distributions rather than their frequency functions, and the next step therefore is to obtain with the aid of these moments approximate expressions for the desired frequency functions. In this connection it should be borne in mind that we are not concerned with questions regarding the probable errors of the moments which we are employing, since the moments computed for the distributions of samples are necessarily exact, and their probable errors are therefore zero. For

[1]See Annals, Vol. I, page 104.

this reason arguments tending to limit the number of terms that may be employed in either a Gram-Charlier series, or in the denominator of Pearson's differential equation are not to the point so far as our illustrations are concerned. These remarks hold even for the third example, since if the moments of the parent population are as given, then the moments of the distribution of samples may be determined with any desired degree of accuracy.

Since it is evident that the solution of our problems now depends upon our obtaining approximate expressions for these distributions whose moments are known, we shall at this point develop a general method of representing discrete distributions which is essentially due to the researches of Charlier. Although the results that we shall obtain are practically those that have also been obtained by Gram, Edgeworth and others, the method that we shall employ is that used by Charlier in "Die Strenge Form des Bernoullischen Theorems."

Let $f(x)$ be the frequency function for a discrete distribution ranging from $x = l_1$ to $x = l_2$. If the ordinates be equidistant at intervals of h, the total frequency of the distribution is

$$(3)\quad N = f(l_1) + f(l_1+h) + \cdots + f(x_0-h) + f(x_0) + f(x_0+h) + \cdots + f(l_2) = \sum_{x=l_1}^{l_2} f(x).$$

where our interest is focused on a typical ordinate at $x = x_0$. If we now set up the function

$$\sum_{x=l_1}^{l_2} f(x)\cdot e^{x\omega i} = f(x_0) + f(x_0+h)\cdot e^{(x_0+h)\omega i} + \cdots + f(l_2)e^{l_2\omega i} + f(x_0-h)e^{(x_0-h)\omega i} + \cdots + f(l_1)e^{l_1\omega i}$$

where $i = \sqrt{-1}$, and multiply each side by $e^{-x_0 \omega i}$ so that

$$e^{-x_0\omega i}\sum_{x=l_1}^{l_2} f(x)e^{x\omega i} = f(x_0) + f(x_0+h)\cdot e^{h\omega i} + f(x_0+2h)\cdot e^{2h\omega i} + \cdots$$

$$+ f(l_2)\cdot e^{(l_2-x_0)\omega i} + f(x_0-h)\cdot e^{-h\omega i} + f(x_0-2h)\cdot e^{-2h\omega i} + \cdots + f(l_1)e^{(l_1-x_0)\omega i}$$

we obtain by integrating both members with respect to ω between the limits $\omega = \frac{-\pi}{h}$ and $\omega = \frac{\pi}{h}$

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-x_0\omega i}\left\{\sum_{x=l_1}^{l_2} f(x)\cdot e^{x\omega i}\right\}d\omega = \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} f(x_0)\, d\omega,$$

since the integral of all other terms of the right hand member will vanish as follows:

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} f(x_0+mh)\cdot e^{mh\omega i} d\omega = f(x_0+mh)\cdot$$

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \left[\cos mh\omega + i \sin mh\omega\right] d\omega = 0$$

(m is an integer.)

It follows therefore that

$$(4) \qquad f(x_0) = \frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-x_0\omega i}\left\{\sum_{x=l_1}^{l_2} f(x)\cdot e^{x\omega i}\right\}d\omega$$

Moreover, since

$$e^{-a\omega i}+e^{-(a+h)\omega i}+\cdots+e^{-b\omega i}=\frac{e^{-(b+h)\omega i}-e^{-a\omega i}}{e^{-h\omega i}-1}$$

we see that the sum of all the consecutive frequencies from $x=a$ to $x=b$ may be expressed as the definite integral

$$(5)\quad \sum_{x=a}^{b} f(x)=\frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \frac{e^{-(b+h)\omega i}-e^{-a\omega i}}{e^{-h\omega i}-1}\left\{\sum_{x=l_1}^{l_2} f(x)\cdot e^{x\omega i}\right\}d\omega$$

The changing of the order of integration is permitted since the limits are all finite.

Ordinarily frequency distributions are expressed as developments of the integral (4), and the sums of consecutive frequencies obtained by applying the Euler-Maclaurin Sum-Formula to these results. It seems at first sight that it might be well to place a little more emphasis upon the evaluation of (5), since this as it stands affords an exact expression for the sum of any group of consecutive frequencies. For the case of continuous variates we need only permit h to approach zero, replace the sign of summation by the sign of integration, etc., and after justifying the change in the order of integration for the resulting infinite limits obtain

$$(6)\int_a^b f(x)dx=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-b\omega i}-e^{-a\omega i}}{-\omega i}\cdot\left\{\int_{l_1}^{l_2} f(x)\cdot e^{x\omega i}dx\right\}d\omega$$

We shall now attempt to evaluate the definite integral (4). Let us first observe that the quantity within the parenthesis is a function of ω , since the finite integration with respect to x and the subsequent replacing of x by the limits will cause this distribution variable to disappear.

For reasons which will develop later, let us write

$$\sum_{x=\ell_1}^{\ell_2} f(x)e^{x\omega i} = e^{b_1(\omega i)+b_2\frac{(\omega i)^2}{2!}} \sum_{x=\ell_1}^{\ell_2} f(x)\cdot e^{(x-b_1)\omega i - \frac{b_2(\omega i)^2}{2!}} .$$

If in Leibnitz' formula

$$D^n u.v = u\cdot D^n v + \binom{n}{1} Du \cdot D^{n-1} v + \binom{n}{2} D^2 u \cdot D^{n-2} v + \cdots$$

we place $u = e^{\frac{bz^2}{2}}$ and $v = e^{az}$, and note that

$$D^{2n+1} e^{\frac{bz^2}{2}}\Big]_{z=0} = 0$$

$$D^{2n} e^{\frac{bz^2}{2}}\Big]_{z=0} = \frac{(2n)!}{2^n \cdot n!} b^n$$

then

$$(7)\quad D^n e^{az+\frac{bz^2}{2}}\Big]_{z=0} = a^n + \frac{n^{(2)}}{2\cdot 1!} a^{n-2} b + \frac{n^{(4)}}{2^2\cdot 2!} a^{n-4} b^2 + \frac{n^{(6)}}{2^3\cdot 3!} a^{n-6} b^3 + \cdots$$

where $n^{(i)} = n(n-1)(n-2) \cdot \cdot \cdot \cdot \cdot$ to i factors.

Thus we may write

$$\sum_{x=\ell_1}^{\ell_2} f(x) e^{(x-b_1)\omega i - b_2 \frac{(\omega i)^2}{2}} = N\left[c_0 + c_1(\omega i) + c_2 \frac{(\omega i)^2}{2!} + c_3 \frac{(\omega i)^3}{3!} + \cdots\right]$$

and employing the notation

$$\sum_{x=\ell_1}^{\ell_2} (x-b_1)^n f(x) = N\mu'_n$$

we obtain from (7)

$$(8)\quad \begin{cases} c_0 = 1 \\ c_1 = \mu'_1 \\ c_2 = \mu'_2 - b_2 \\ c_3 = \mu'_3 - 3b_2\mu'_1 \\ c_4 = \mu'_4 - 6b_2\mu'_2 + 3b_2^2 \\ \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ c_n = \mu'_n - \frac{n^{(2)}}{2\cdot 1!} b_2\mu'_{n-2} + \frac{n^{(4)}}{2^2\cdot 2!} b_2^2\mu'_{n-4} - \frac{n^{(6)}}{2^3\cdot 3!} b_2^3\mu'_{n-6} + \cdots \end{cases}$$

Formula (4) may therefore be written, dropping the subscript on x_0

$$(9)\qquad f(x) = N\frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-(x-b_1)\omega i - b_2\frac{(\omega i)^2}{2}} \cdot$$

$$\left[1 + c_1(\omega i) + c_2\frac{(\omega i)^2}{2} + \cdots\right] d\omega$$

Placing

$$(10) \qquad \Theta(x) = \frac{1}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-(x-b_1)\omega i - \frac{b_2\omega^2}{2}}\, d\omega$$

it follows that the nth derivative with respect to x is

$$(11) \qquad \Theta^{(n)}(x) = \frac{1}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} (-\omega i)^n \cdot e^{-(x-b_1)\omega i - \frac{b_2\omega^2}{2}}\, d\omega\,;$$

so finally

$$(12) \qquad f(x) = N\cdot h\left[\Theta(x) - \frac{c_1}{1!}\Theta^{(1)}(x) + \frac{c_2}{2!}\Theta^{(2)}(x) - \frac{c_3}{3!}\Theta^{(3)}(x) + \cdots\right]$$

Let us now investigate the function $\Theta(x)$.

$$\Theta(x) = \frac{1}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-b_2\omega^2/2}\cdot\Big[\cos(x-b_1)\omega$$

$$-\, i\sin(x-b_1)\omega\Big]\, d\omega$$

$$= \frac{1}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-b_2\omega^2/2}\cos(x-b_1)\,\omega\, d\omega$$

[since $e^{-b_2\omega^2/2}\sin(x-b_1)\omega$ is an odd function of ω]

$$=\frac{1}{\pi}\int_0^\infty e^{-b_2\omega^2/2}\cos(x-b_1)\omega\, d\omega$$

$$-\frac{1}{\pi}\int_{\frac{\pi}{h}}^\infty e^{-b_2\omega^2/2}\cos(x-b_1)\omega\, d\omega$$

$$=\frac{1}{\sqrt{2\pi b_2}}e^{-\frac{(x-b_1)^2}{2b_2}}-\frac{1}{\pi}\int_{\frac{\pi}{h}}^\infty e^{-b_2\omega^2/2}\cos(x-b_1)\omega\, d\omega$$

$$=\phi(x)-R_0.$$

$$\left[\int_0^\infty e^{-a^2x^2}\cos mx\,dx=\frac{\sqrt{\pi}}{a}\,e^{-m^2/4a^2}\right]$$

Likewise we may write

$$\theta^{(n)}(x)=\phi^{(n)}(x)-R_n,\qquad R_n<\frac{1}{\pi}\int_{\frac{\pi}{h}}^\infty \omega^n e^{-b_2\omega^2/2}\,d\omega$$

By successive integration by parts it can be shown that

$$\int x^n e^{-\frac{x^2}{2}}dx=-e^{-\frac{x^2}{2}}\Big\{x^{n-1}+(n-1)x^{n-3}$$

$$(13)\qquad +(n-1)(n-3)x^{n-5}+\cdots+(n-1)(n-3)\cdots(n-2i+3)x^{n-2i+1}\Big\}+R_i,$$

$$R_i=(n-1)(n-3)\cdots(n-2i+1)\int x^{n-2i}e^{-\frac{x^2}{2}}dx$$

so we have that

$$(14)\quad R_n < \frac{1}{b_2}\left(\frac{\pi}{h}\right)^{n-1} e^{-\frac{b_2\left(\frac{\pi}{h}\right)^2}{2}}\left[1+\frac{n-1}{b_2}\left(\frac{h}{\pi}\right)^2+\frac{(n-1)(n-3)}{b_2^2}\left(\frac{h}{\pi}\right)^4+\cdots\right]$$

So far we have said nothing concerning the values of the parameters b_1 and b_2. Referring to formula (8) it is seen that if the origin of x be taken at the mean of the distribution in question, and b_2 equal the second moment about the mean of this distribution, $c_1 = c_2 = 0$, and consequently if the values of R_n may be neglected, the equation of the distribution expressed in standard units becomes

$$(15)\quad f(x)=N\frac{h}{\sigma}\left\{\phi(t)-\frac{A_3}{3!}\phi^{(3)}(t)+\frac{A_4}{4!}\phi^{(4)}(t)-\frac{A_5}{5!}\phi^{(5)}(t)+\cdots\right\}$$

where $t=\frac{x-b_1}{\sqrt{b_2}}=\frac{x-M}{\sigma}$, $\phi(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$, and

$$(16)\quad \begin{cases} A_3=\alpha_3 \\ A_4=\alpha_4-3 \\ A_5=\alpha_5-10\alpha_3 \\ A_6=\alpha_6-15\alpha_4+30 \\ \quad\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ A_n=\alpha_n-\frac{n^{(2)}}{2\cdot 1!}\alpha_{n-2}+\frac{n^{(4)}}{2^2\cdot 2!}\alpha_{n-4}-\frac{n^{(6)}}{2^3\cdot 3!}\alpha_{n-6}+\cdots \end{cases}$$

By employing the Euler-Maclaurin Sum-Formula we can write

$$
\begin{aligned}
&f(a)+f(a+h)+f(a+2h)+\cdots+f(b-h)+f(b) \\
&= N\left[\int \phi(t)\,dt - A'_0\phi(t) + A'_1\overset{(1)}{\phi}(t) - A'_2\overset{(2)}{\phi}(t) + A'_3\overset{(3)}{\phi}(t)\cdots\right]_{\frac{a-M}{\sigma}}^{\frac{b+h-M}{\sigma}}.
\end{aligned} \tag{17}
$$

where

$$
\left\{
\begin{aligned}
A'_0 &= \frac{h}{2\sigma} \\
A'_1 &= \frac{h^2}{12\sigma^2} \\
A'_2 &= \frac{\alpha_3}{6} \\
A'_3 &= \frac{\alpha_4-3}{24} + \frac{h}{\sigma}\cdot\frac{\alpha_3}{12} - \frac{h^4}{720\sigma^4} \\
A'_4 &= \frac{\alpha_5-10\alpha_3}{120} + \frac{h}{\sigma}\ \frac{\alpha_4-3}{48} + \frac{h^2}{\sigma^2}\cdot\frac{\alpha_3}{72} \\
A'_5 &= \frac{\alpha_6-15\alpha_4+30}{720} + \frac{h}{\sigma}\ \frac{\alpha_5-10\alpha_3}{240} + \frac{h^2}{\sigma^2}\cdot\frac{\alpha_4-3}{288} + \frac{h_6}{30240\sigma^6}
\end{aligned}
\right. \tag{18}
$$

In some cases it may be more convenient to employ a mean and a standard deviation of the generating function that differs somewhat from that of the distribution for which the representation is desired. In this event the coefficients of the first and second derivatives in (15) will not vanish. However, the extra effort

expended in increasing the number of significant terms may be more than offset by the fact that a rather arbitrary choice in the values of b_1 and b_2 may result in simpler values for

$$t = \frac{x - b_1}{\sqrt{b_2}}$$

which in turn may occasionally eliminate difficult interpolations when dealing with tabulations of the generating function and its derivatives.

Formulae (17) and (18) may be regarded as a sort of apology for the fact that the definite integral of formula (5) has never been developed. The need of a satisfactory expression for the sum of any number of consecutive variates is indeed acute.

By permitting h in the foregoing theory to approach zero, one can obtain corresponding formulae for the ordinates and areas of distributions of continuous variates. However, it should be noted that for this case the limits for the integrals in the vicinity of formula (4) are now

$$\lim_{h \to 0} \frac{\pi}{h} = \infty$$

and consequently the changing of the order of integration must be justified.

In conclusion we may state:

I. Answers to problems of statistical sampling are usually expressed as finite or infinitesimal integrals under a function whose moments *only* are known. If known, the function is generally of but little value.

II. It is necessary to approximate the desired integrals by employing frequency functions.

III. Present methods are unsatisfactory from the point of view that remainder or limit of error terms are not available. The χ^2 test, though helpful, does not meet the issue in question.

H. C. Carver.

NOTE ON THE DISTRIBUTION OF MEANS OF SAMPLES OF N DRAWN FROM A TYPE A POPULATION

By

CECIL C. CRAIG
National Research Fellow

Recently in this journal, Dr. George A. Baker has found "the distribution of the means of samples drawn at random from a population represented by a Gram-Charlier series."[1] It is the purpose of this note to call attention to the fact that by the use of the semi-invariant notation Dr. Baker's results may be reached in very many fewer steps.

Let the parent population be represented by

$$(1) \quad f(x) = \phi(x)\left[1 + \frac{a_3}{\sigma_x^3} H_3\left(\frac{x}{\sigma_x}\right) + \frac{a_4}{\sigma_x^4} H_4\left(\frac{x}{\sigma_x}\right) + \cdots + \frac{a_k}{\sigma_x^k} H_k\left(\frac{x}{\sigma_x}\right)\right]$$

in which

$$(2) \quad \phi(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma_x^2}}$$

[1]Vol. 1, No. 3 (Aug., 1930), pp. 199-204.

the origin for x being chosen at the mean, and

$$H_k(t)e^{-\frac{t^2}{2}} = D_t^k\left(e^{-\frac{t^2}{2}}\right). \tag{3}$$

We shall first find the distribution function of $z = x_1 + x_2 + \cdots + x_N$ in which x_i, $i = 1, 2, \cdots N$, has the frequency function $f(x)$. Let us assume the frequency function of z is given by

$$F(z) = \phi(z)\left[1 + \frac{A_3}{\sigma_z^3} H_3\left(\frac{z}{\sigma_z}\right) + \frac{A_4}{\sigma_z^4} H_4\left(\frac{z}{\sigma_z}\right) + \cdots + \frac{A_l}{\sigma_z^l} H_l\left(\frac{z}{\sigma_z}\right)\right] \tag{4}$$

Then the semi-invariants of $f(x)$, $\lambda_1, \lambda_2, \cdots \lambda_k$ are defined by the formal identity in t:

$$e^{\lambda_1 t + \frac{1}{2!}\lambda_2 t^2 + \frac{1}{3!}\lambda_3 t^3 + \cdots} = \int_{-\infty}^{\infty} dx\, f(x) e^{xt} \quad (\lambda_1 = 0 \text{ in this case}) \tag{5}$$

and on integration, using (3), we get at once on the right:

$$e^{\lambda_2 \frac{t^2}{2}}\left[1 - a_3 t^3 + a_4 t^4 + \cdots (-1)^k a_k t^k\right]$$

Similarly for the semi-invariants $L_1, L_2, L_3 \cdots$ of $F(z)$ we have

$$e^{L_1 t + \frac{1}{2!}L_2 t^2 + \frac{1}{3!}L_3 t^3 + \cdots} = e^{L_2 \frac{t^2}{2}}\left[1 - A_3 t^3 + A_4 t^4 - \cdots + (-1)^l A_l t^l + \cdots\right] \tag{6}$$

But because of the well-known fact that $L_r = N\lambda_r$ this gives

$$1-A_3t^3+A_4t^4-\cdots-(-1)^l A_l t^l$$
$$=\left[1-a_3t^3+a_4t^4-\cdots(-1)^k a_k t^k\right]^N$$

an identity in t. Thus

$$(7)\quad A_r=\sum\frac{N!}{\nu_3!\,\nu_4!\cdots\nu_k!\,(N-\nu_3-\nu_4\cdots\nu_k)!}\,a_3^{\nu_3}a_4^{\nu_4}\cdots a_k^{\nu_k}$$

the summation including all terms for which

$$3\nu_3+4\nu_4+\cdots+k\nu_k=r$$

Remembering that $\sigma_z=\sqrt{L_2}=\sqrt{N}\,\sigma_x$, we have on substitution in (4) the expression for $F(z)$ since only a finite number of A_r's (depending on N) are different from zero.

To get the distribution of $z'=\dfrac{x_1+x_2+\cdots+x_N}{N}$ only involves the appropriate change of unit.

Cecil C. Craig

Stanford University.